Advanced Financial Models

Michael R. Tehranchi

Contents

This course is about models of financial markets, with an emphasis on the pricing and hedging of contingent claims within such models. Our starting point is the self-evident observation:

The future is uncertain.

Indeed, anyone with even a passing acquaintance with finance knows that it is impossible to predict with absolute certainty how the the price of an asset will fluctuate. Therefore, the proper language to formulate the models that we will study is the language of probability theory. An attempt is made to keep this course self-contained, but you should be familiar with the basics of the theory, including knowing the definition and key properties of the following concepts: random variable, expected value, variance, conditional probability/expectation, independence, Gaussian (normal) distribution, etc. Familarity with measure theoretical probability is helpful, though a crashcourse on probability theory is given in an appendix.

Please send all comments (including small typos and major blunders) to the author at `m.tehranchi@statslab.cam.ac.uk`.

CHAPTER 1

One-period models

The models we will encounter will be of form $(S_t^0, \dots, S_t^d)_{t\in\mathbb{T}}$ where S_t^i will model the price of a financial asset (stock, bond, etc.) at time $t \in \mathbb{T}$. In this course, the index set $\mathbb{T}$ will be one of the three sets

- $\{0, 1\}$ for single period models,
- $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$ when time is discrete, and
- $\mathbb{R}_+ = [0, \infty)$ when time is continuous.

As simple as it seems, much of the financial aspects of this course already appear in one-period models where $\mathbb{T} = \{0, 1\}$. It is, therefore, appropriate to devote a significant portion of the course to this important special case.

1. The set-up

In this section we describe our first model. As we shall see later, this set-up captures most of the essential features of the general discrete-time model.

The market consists of $d + 1$ assets. We assume now, and we will leave it as a standing assumption throughout the course, that an investor in this market can buy or sell any number (even fractional) of shares of each asset without affecting the price. Of course, in the real world, investors face constraints when short-selling assets, and large buy or sell orders tend to move the market prices, but we ignore these issues. We also assume that there is no bid/ask spread – the buying price is the same as the selling price.

We model the price of the assets, labelled $0, 1, \dots, d$, at time $t \in \{0, 1\}$ by the random variable S_t^i. We will denote the collection of prices by the $d + 1$-dimensional column vector

$$\bar{S}_t = (S_t^0, \dots, S_t^d).$$

Remark. You might wonder at this stage why we have chosen the notation such that there are $d + 1$ assets, rather than simply d assets. The reason is that in the sequel, one of the assets, usually asset 0, plays a distinguished role. Often, but not always, we take asset 0 to be ordinary cash, so that $S_0^0 = S_1^0 = 1$. Or, we can model the cost of borrowing cash by introducing an interest rate $r \geq 0$ so that $S_0^0 = 1$ and $S_1^0 = 1 + r$. Then the market consists of one distinguished asset and d other assets, for a total of $d + 1$. Also, the overbar in the notation $\bar{S}_t$ will become apparent as this chapter proceeds.

Our first assumption on the stochastic process $\bar{S} = (\bar{S}_t)_{t\in\{0,1\}}$ is that the time 0 prices of the assets are known at time 0. Formally, we have the following assumption (which will be generalized in the multi-period models):

Assumption. The random variables $S_0^0, \dots, S_0^d$ are constants – that is, not random.

Now given this market model, we introduce an investor. Since there is only one period, the investor only chooses his investment portfolio once, at time 0. We use the following notation:

- $\pi^i \in \mathbb{R}$ denotes the (non-random) number of shares of asset i, for $i = 0, \dots, d$, and let
$$\bar{\pi} = (\pi^0, \dots, \pi^d)$$
denote the vector of portfolio weights.
- $X_t(\bar{\pi})$ denotes the investor's wealth at time $t \in \{0, 1\}$.

Remark. We will allow π^i to be either positive, negative, or zero with the interpretation that if $\pi^i > 0$ the investor is 'long' asset i and if $\pi^i < 0$ the investor is 'short' the asset. In particular, if $\pi^i < 0$, then the investor has borrowed $|\pi^i|$ units of asset i to pay back at time 1. Again notice that we do not demand that the π^i are integers.

We assume that the investor's wealth and porfolio is connected by the following relationships:

$$\boxed{\begin{array}{rclr} X_0(\bar{\pi}) & = & \bar{\pi} \cdot \bar{S}_0 & \text{the budget constraint} \\ X_1(\bar{\pi}) & = & \bar{\pi} \cdot \bar{S}_1 & \text{the self-financing condition} \end{array}}$$

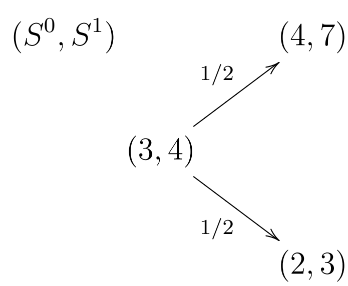
where the $a \cdot b = \sum_{i=0}^d a^i b^i$ is the usual Euclidean inner (or dot) product in $\mathbb{R}^{d+1}$.

Remark. The budget constraint simply says that the investor's initial wealth is the time-0 total value of his holdings, and the self-financing condition says that changes in his wealth between time 0 and 1 are due only to changes in the asset prices – he does not consume or have any other source of income.

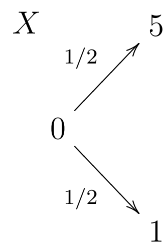
2. Arbitrage and the first fundamental theorem of asset pricing

Now that we have our market model and we've introduced an investor into this market, our first challenge is to find out how to invest optimally. An investor's dream is to find a portfolio $\bar{\pi}$ that costs nothing to buy at time 0, but that always has non-negative (and sometimes positive) value at time 1.

Example. Consider a market with two assets with prices given by



(The above diagram should be read $S_0^0 = 3$, $\mathbb{P}(S_1^1 = 7) = 1/2$, etc.) Consider the portfolio $\pi^0 = -4, \pi^1 = 3$. It costs zero to buy at time 0, but the time-1 wealth is strictly positive in all states of the world:



The example above leads us to our first definition:

Definition. An *arbitrage* is a portfolio $\bar{\pi} \in \mathbb{R}^{d+1}$ such that

- $X_0(\bar{\pi}) \leq 0$,
- $X_1(\bar{\pi}) \geq 0$ almost surely, and
- either $X_0 < 0$ or $\mathbb{P}(X_1 > 0) > 0$ (or both).

At this point, you should check that the portfolio $\bar{\pi} = (-3, 2)$ is also an arbitrage in the example above.

Markets with many arbitrage opportunities would be nice–we all would be a lot richer. But for the sake of building realistic models, we usually assume that markets are free of arbitrages. Notice that

a market has *no arbitrage*	if and only if $\bar{\pi} \cdot \bar{S}_0 \leq 0$ and $\bar{\pi} \cdot \bar{S}_1 \geq 0$ a.s. $\Rightarrow \bar{\pi} \cdot \bar{S}_0 = 0$ and $\bar{\pi} \cdot \bar{S}_1 = 0$ a.s.

In this section we find a mathematical classification of such market models. Before we begin, we need some vocabulary.

Definition. A *pricing kernel* (or *state price density*) for a market model $\bar{S}$ is a positive random variable ρ such that

$$\mathbb{E}(\rho |S_1^i|) < \infty, \text{ and } \mathbb{E}(\rho S_1^i) = S_0^i.$$

for all $i = 0, \ldots, d$.

Now we come to first theorem of the course, and one of the most important theorems in financial mathematics. It is no surprise that it is often called the first fundamental theorem of asset pricing. The following proof is from Doug Kennedy's lecture notes.

Theorem (First fundamental theorem of asset pricing). *A market model has no arbitrage if and only if there exists a pricing kernel.*

Proof. First we prove that if there exists a pricing kernel, then there is no arbitrage. Letting $X_t = \bar{\pi} \cdot \bar{S}_t$, we have by the definition of pricing kernel and the linearity of expectations

$$\begin{aligned} \mathbb{E}(\rho X_1) &= \bar{\pi} \cdot \mathbb{E}(\rho \bar{S}_1) \\ &= \bar{\pi} \cdot \bar{S}_0 \\ &= X_0. \end{aligned}$$

Now suppose $X_0 \leq 0$ and $X_1 \geq 0$ almost surely. Since $\rho > 0$, we have $\rho X_1 \geq 0$ and hence

$$0 \geq X_0 = \mathbb{E}(\rho X_1) \geq 0$$

so that $X_0 = 0 = \mathbb{E}(\rho X_1)$. We also see that $\rho X_1 = 0$ a.s. (Recall the pigeonhole principle: if $Y \geq 0$ a.s and $\mathbb{E}(Y) = 0$ then $Y = 0$ a.s.) Again, since $\rho \neq 0$ a.s., we conclude that $X_1 = 0$ a.s., and hence there is no arbitrage.

We now show that if there does not exist a pricing kernel, then there must be an arbitrage opportunity. Consider the set $C \subseteq \mathbb{R}^{d+1}$ defined by

$$C = \{\mathbb{E}(\rho \bar{S}_1) : \rho > 0 \text{ a.s. and } \mathbb{E}(\rho|\bar{S}_1|) < \infty\}.$$

The set C is not empty, since ρ_0 defined by $\rho_0 = e^{-|\bar{S}_1|}$ certainly satisfies $\rho_0 > 0$ a.s. and $\mathbb{E}(\rho_0|\bar{S}_1|) < \infty$. Also, it is easy to see that C is convex.

If $\bar{S}_0$ is contained in C, there would exist a pricing kernel. So, we suppose that $\bar{S}_0$ is not an element of C. By the separating hyperplane theorem (stated and proved below) there exists a vector $\bar{\pi}$ such that

$$\bar{\pi} \cdot (y - \bar{S}_0) \geq 0$$

for all $y \in C$, and such that the above inequality is strict for at least one $y \in C$. Letting $\bar{\pi} \cdot \bar{S}_t = X_t$ the above inequality translates to

$$(*) \qquad X_0 \leq \mathbb{E}(\rho X_1)$$

for all feasible ρ, with strict inequality for at least one ρ.

First, by letting $\rho = \epsilon \rho_0$ with $\epsilon > 0$ in the above inequality (*), we have

$$X_0 \leq \epsilon \mathbb{E}(\rho_0 X_1) \to 0$$

as $\epsilon \downarrow 0$, so we can conclude $X_0 \leq 0$.

Next, let $\rho = \rho_0(M \mathbb{1}_{\{X_1 < 0\}} + 1)$ for $M > 0$ so that inequality (*) becomes

$$X_0 \leq M \mathbb{E}(\rho_0 X_1 \mathbb{1}_{\{X_1 < 0\}}) + \mathbb{E}(\rho_0 X_1).$$

Dividing by M and sending $M \uparrow \infty$ yields

$$\mathbb{E}(\rho_0 X_1 \mathbb{1}_{\{X_1 < 0\}}) \geq 0.$$

Since $\rho_0 X_1 \mathbb{1}_{\{X_1 < 0\}} \leq 0$ a.s. we must conclude $X_1 \geq 0$ a.s.

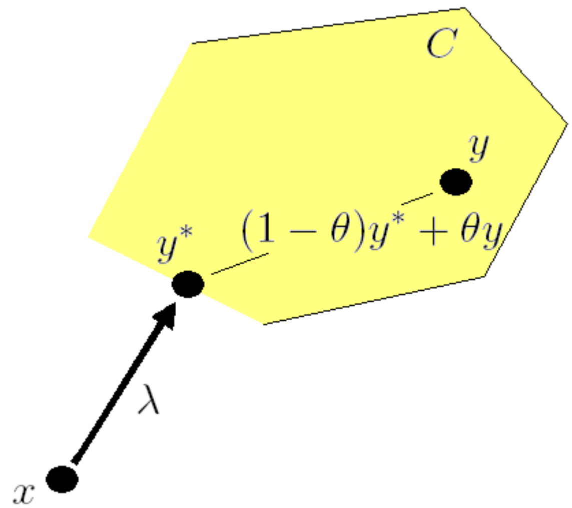
We have just shown $X_0 \leq 0$ and $X_1 \geq 0$ a.s. The separating hyperplane theorem says that there exists at least one ρ such that

$$X_0 < \mathbb{E}(\rho X_1),$$

so we must conclude that either $X_0 < 0$ or $\mathbb{P}(X_1 > 0) > 0$ (or both). Hence $\bar{\pi}$ is our desired arbitrage. □

2.1. A separating hyperplane theorem*. In this optional subsection, a version of the separating hyperplane theorem is stated and proved. There are many versions of this theorem, but we give only the version needed to prove the first fundamental theorem of asset pricing in one-period.

Theorem (Separating/supporting hyperplane theorem). *Let $C \subset \mathbb{R}^N$ be convex and $x \in \mathbb{R}^N$ not contained in C. Then there exists a $\lambda \in \mathbb{R}^N$ such that $\lambda \cdot (y - x) \geq 0$ for all $y \in C$, where the inequality is strict for at least one $y \in C$.*



Case 1. Separating hyperplane

Proof. Case 1: Separating hyperplane. The point x is not in the closure $\bar{C}$ of C. Let $y^* \in \bar{C}$ be the point in $\bar{C}$ closest to x. Assuming for the moment the existence of this point, let $\lambda = y^* - x$. Fix a point $y \in C$ and $0 < \theta < 1$, and note that the point $(1-\theta)y^* + \theta y$ is in $\bar{C}$ since $\bar{C}$ is convex. Then

$$\begin{aligned} 0 &= |y^* - x|^2 - \lambda^2 \\ &\leq |(1-\theta)y^* + \theta y - x|^2 - \lambda^2 \\ &= |\theta(y - y^*) + \lambda|^2 - \lambda^2 \\ &= \theta^2 |y - y^*|^2 + 2\theta(y - y^*) \cdot \lambda. \end{aligned}$$

By first dividing by θ and then taking the limit as $\theta \downarrow 0$ in the above inequality, we conclude $(y - y^*) \cdot \lambda \geq 0$. Hence

$$(y - x) \cdot \lambda = (y - y^*) \cdot \lambda + |\lambda|^2 > 0$$

as desired.

Now, we establish the existence of y^*: Let $d = \inf_{y \in C} |y - x| > 0$ and let y_n be a sequence in C such that $|y_n - x| \to d$. Applying the parallelogram law $|a+b|^2 + |a-b|^2 = 2|a|^2 + 2|b|^2$ we have

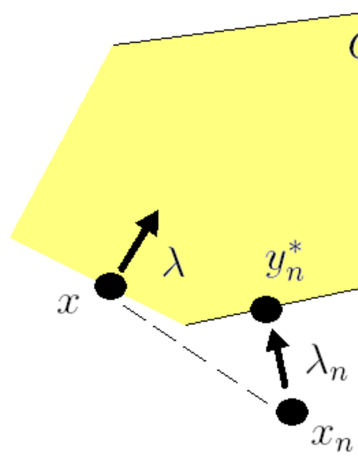
$$\begin{aligned} |y_m - y_n|^2 &= 2|y_m - x|^2 + 2|y_n - x|^2 - 4\left|\tfrac{1}{2}(y_m + y_n) - x\right|^2 \\ &\leq 2|y_m - x|^2 + 2|y_n - x|^2 - 4d^2 \to 0 \end{aligned}$$

as $m, n \to \infty$, where we have used the convexity of C to assert that $\frac{1}{2}(y_m + y_n) \in C$ and hence $\left|\frac{1}{2}(y_m + y_n) - x\right| \geq d$. We have established that the sequence $(y_n)_n$ is Cauchy, and hence converges to some point $y^* \in \bar{C}$ as claimed.

Case 2: Supporting hyperplane. The point x is in the closure $\bar{C}$ of C. Define a subspace of $\mathbb{R}^N$ by

$$S = \operatorname{span}\{y - x : y \in C\}.$$

Let $(x_n)_n$ be a sequence in the complement of $\bar{C}$ such that $x_n - x$ is in S and $x_n \to x$. As



Case 2. Supporting hyperplane

in case 1, we can find the point $y_n^* \in \bar{C}$ closest to x_n. Let

$$\lambda_n = \frac{y_n^* - x_n}{|y_n^* - x_n|}.$$

Since $(\lambda_n)_n$ is bounded, there exists a convergent subsequence, still denoted by $(\lambda_n)_n$, with limit $\lambda \in S$. Then

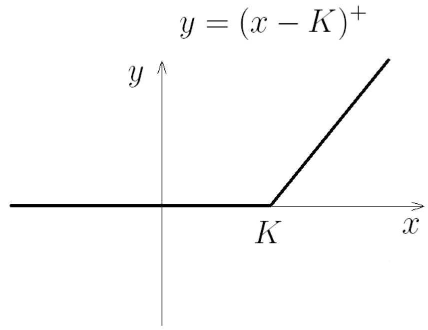
$$\begin{aligned} \lambda \cdot (y - x) &= \lim_n \lambda_n \cdot (y - x) \\ &= \lim_n \lambda_n \cdot (y - x_n) + \lambda_n \cdot (x_n - x) \\ &\geq 0. \end{aligned}$$

Finally, since $\lambda \in S$ and $\lambda \neq 0$, we can conclude that there exists $y \in C$ such that $\lambda \cdot (y - x) \neq 0$, as desired. □

3. Contingent claims and no-arbitrage pricing

A contingent claim is any cash payment at time 1 where the size of the payment is contingent on the realized prices of other assets, etc. In other words, for us, the payout of a contingent claim is modelled by a random variable on our probability space $(\Omega, \mathcal{F}, \mathbb{P})$. One of the major triumphs of modern finance is the no-arbitrage pricing of contingent claims.

Example (Call option). A call option gives the owner of the option the right, but not the obligation, to buy a given stock at time 1 at some fixed price K, called the *strike* of the option. Let S_1 denote the price of the stock at time 1. There are two cases: If $K \geq S_1$, then the option is worthless to the owner since there is no point paying a price above the market price for the underlying stock. On the other hand, if $K < S_1$, then the owner of the option can buy the stock for the price K from the counterparty and immediately sell the stock for the price S_1 to the market, realizing a profit of $S_1 - K$. Hence, the payout of the call option is $(S_1 - K)^+$, where $a^+ = \max\{a, 0\}$ as usual. The 'hockey-stick' graph of the function $g(x) = (x - K)^+$ is below.



We now study what the no-arbitrage principle has to say about the pricing of contingent claims. We start off with a given market model of $d + 1$ assets with prices $\bar{S}$, and suppose this market is free of arbitrage. Introduce a contingent claim with time-1 payout of ξ_1. Our goal is to find the time-0 price ξ_0 such that the augmented market with prices $(\bar{S}, \xi)$ is still free of arbitrage. There is a useful class of contingent claims for which no-arbitrage give a unique price.

Definition. A contingent claim with payout ξ_1 is an *attainable* (or *replicable*) claim if and only if there exists a portfolio $\bar{\pi} \in \mathbb{R}^{d+1}$ such that

$$\xi_1 = \bar{\pi} \cdot \bar{S}_1.$$

This set of attainable claims can be characterized:

Theorem (Characterization of attainable claims). *Given an arbitrage free market model $\bar{S}$, the following are equivalent:*

(1) *A contingent claim with payout ξ_1 is attainable.*
(2) *There exists a unique initial price ξ_0 such that the augmented market $(\bar{S}, \xi)$ has no arbitrage.*
(3) *There is a constant ξ_0 such that $\xi_0 = \mathbb{E}(\rho\xi_1)$ for all pricing kernels ρ such that $\mathbb{E}(\rho|\xi_1|) < \infty$.*

Remark. The conclusion of the theorem is very important, so it's worth reprasing it for emphasis. The theorem says that a contingent claim has a unique no-arbitrage price if and only if its payout can be replicated by trading in the original assets $0, \ldots, d$. Furthermore, the time-0 price of such a claim, which is just the time-0 price of this replicating portfolio, can be found by computing the expected value of the payout times a pricing kernel.

Following proof is a bit redundant but may help with building intuition.

Proof. (1) $\Rightarrow$ (2) Suppose the claim is attainable, so that $\xi_1 = \bar{\pi} \cdot \bar{S}_1$ for some portfolio $\bar{\pi}$. We will show that if the augmented market $(\bar{S}, \xi)$ has no-arbitrage, then $\xi_0 = \bar{\pi} \cdot \bar{S}_0$. So, by symmetry, it's enough to show that if $\xi_0 < \bar{\pi} \cdot \bar{S}_0$ then there exists an arbitrage.

Consider a portfolio $\phi = (\phi^0, \ldots, \phi^d, \phi^{d+1})$ where the first $d + 1$ entries correspond to the number of shares of the assets in the underlying market, and the $d + 2$-th entry is for the contingent claim. Let

$$\phi^i = -\pi^i, \text{ for } i = 0, \ldots, d, \text{ and } \phi^{d+1} = 1.$$

Then

$$X_0(\phi) = -\bar{\pi} \cdot \bar{S}_0 + \xi_0 < 0$$

and

$$X_1(\phi) = -\bar{\pi} \cdot \bar{S}_1 + \xi_1 = 0 \text{ a.s.}$$

Hence ϕ is an arbitrage.

(1) $\Rightarrow$ (3) If ξ_1 is attainable, then $\xi_1 = \bar{\pi} \cdot \bar{S}_1$ for some portfolio $\bar{\pi}$. Then

$$\mathbb{E}(\rho \xi_1) = \bar{\pi} \cdot \bar{S}_0$$

for every pricing kernel ρ, so that $\xi_0 = \bar{\pi} \cdot \bar{S}_0$.

(3) $\Leftrightarrow$ (2) By the first fundamental theorem of asset pricing, the augmented market $(\bar{S}, \xi)$ has no arbitrage if and only if there exists a positive random variable ρ such that

$$\mathbb{E}(\rho \bar{S}_1) = \bar{S}_0 \text{ and } \mathbb{E}(\rho \xi_1) = \xi_0.$$

The first equation above says that ρ is a pricing kernel for the underlying market $\bar{S}$. The second equation says that only there is a unique initial price ξ_0 compatible with no-arbitrage if and only if $\mathbb{E}(\rho \xi_1) = \xi_0$ all pricing kernels ρ.

(3) $\Rightarrow$ (1)* This is the hard direction. Suppose that the sample space Ω has $n + 1$ elements, with $\mathbb{P}\{\omega\} > 0$ for all $\omega \in \Omega$. (The statement is true in full generality, but the proof is left as an exercise on the example sheet.) We can identify a random variable Y with the vector $(Y(\omega_0), \ldots, Y(\omega_n)) \in \mathbb{R}^{n+1}$.

Let ρ_0 be a pricing kernel for $\bar{S}$ and consider the set

$$K = \{\bar{\pi} \cdot \bar{S}_1 : \bar{\pi} \in \mathbb{R}^{d+1}\}$$

of attainable claims. Viewed as a vector subspace of $\mathbb{R}^{n+1}$ of dimension at most $d+1$, it has an orthogonal complement

$$K^\perp = \{\mu : \mathbb{E}(\mu X_1) = 0 \text{ for all } X_1 \in K\}$$

dimension at least $n - d$. (Of course, $K^\perp$ is allowed to be the singleton $\{0\}$.) Pick a vector $\mu \in K^\perp$ and let

$$\rho_\theta = \rho_0 + \theta \mu.$$

Since $\rho_0(\omega) > 0$ for all $\omega \in \Omega$, one can choose a non-zero θ small enough that $\rho_\theta(\omega) > 0$ for all $\omega \in \Omega$. (This is where we have used the assumption that Ω is finite.) Furthermore,

$$\mathbb{E}(\rho_\theta S_1^i) = \mathbb{E}(\rho_0 S_1^i) + \theta \mathbb{E}(\mu S_1^i) = S_0^i.$$

and hence ρ_θ is a pricing kernel as well. Now by assumption

$$\mathbb{E}(\rho_\theta \xi_1) = \mathbb{E}(\rho_0 \xi_0) + \theta \mathbb{E}(\mu \xi_1) = \xi_0.$$

Since $\theta \neq 0$, the vector ξ_1 is orthogonal to μ. But $\mu \in K^\perp$ was arbitrary, we conclude that $\xi_1 \in K^{\perp\perp} = K$ as desired. □

Example. (Put-call parity formula) Suppose we start with a market with three assets with prices $(B_t, S_t, C_t)_{t \in \{0,1\}}$. We assume that asset 0 is a riskless bond with prices $B_0 = 1$ and $B_1 = 1 + r$, that asset 1 is a stock, and that asset 2 is a . a call option on that stock with strike K. Recall that the payout of the option is $C_1 = (S_1 - K)^+$. Suppose that this market is free of arbitrage.

Now we introduce another claim, called a *put* option. A put option gives the owner of the option the right, but not the obligation, to sell the stock for a fixed *strike* price K' at time 1. Using a similar argument as we used for the call option, the payout of a put option is $P_1 = (K' - S_1)^+$.

In the special case when $K = K'$, a miracle occurs, and the put option is attainable in the market (B, S, C). Indeed, we have the identity

$$\begin{aligned} P_1 &= (K - S_1)^+ \\ &= K - S_1 + (S_1 - K)^+ \\ &= \frac{K}{1+r} B_1 - S_1 + C_1, \end{aligned}$$

so the replicating portfolio is $(\pi^0, \pi^1, \pi^2) = \left(\frac{K}{1+r}, -1, 1\right)$. Since the time-0 price of an attainable claim is just the price of the replicating portfolio, we have just derived the famous put-call parity formula

$$\boxed{C_0 - P_0 = S_0 - \frac{1}{1+r} K}$$

Since finding the time-0 price of an attainable claim is easy, we single out the markets for which every claim is attainable:

Definition. An arbitrage-free market is *complete* if and only if every random variable is attainable. A market is *incomplete* otherwise.

We can characterize complete markets:

Theorem (Second Fundamental Theorem of Asset Pricing). *A market model is complete if and only if there exists a unique pricing kernel.*

Proof. First, suppose the market is complete. Let ρ and ρ' be two pricing kernels, and ξ_1 be an arbitrary random variable. Since ξ_1 is attainable by assumption, we have $\mathbb{E}(\rho|\xi_1|) < \infty$ and $\mathbb{E}(\rho'|\xi_1|) < \infty$ and

$$\mathbb{E}(\rho \xi_1) = \xi_0 = \mathbb{E}(\rho' \xi_1)$$

so that

$$\mathbb{E}[(\rho - \rho')\xi_1] = 0.$$

Since the market is complete, every random variable is attainable, so we can let $\xi_1 = \rho - \rho'$ in the above equation. Since the integrand is non-negative, we must conclude $\rho = \rho'$ almost surely, completing the proof of the "only-if" direction.

Now suppose that the pricing kernel ρ is unique, and let ξ_1 be an arbitrary random variable. Assuming $\mathbb{E}(\rho|\xi_1|) < \infty$, we see that $\xi_0 = \mathbb{E}(\rho\xi_1)$ satisfies our characterization of attainable claims, and hence ξ_1 is attainable. We will be done once we show that every ξ_1 has the desired integrability property. All we need do is replace the set C in our proof of the first fundamental theorem with

$$C' = \{\mathbb{E}(\varrho \bar{S}_1) : \varrho > 0 \text{ a.s. and } \mathbb{E}(\varrho|\bar{S}_1|) < \infty, \mathbb{E}(\varrho|\xi_1|) < \infty\}.$$

Following the argument, we see that the no-arbitrage assumption implies the existence of a pricing kernel ρ with $\mathbb{E}(\rho|\xi_1|) < \infty$. But there is only one pricing kernel, so every random variable satisfies the integrability property, concluding the proof of the "if" direction. □

Remark. Suppose that our sample space $\Omega = \{\omega_0, \dots, \omega_n\}$ has $n+1$ points (informally, n sources of randomness). If ρ is a pricing kernel, it must satisfy the $d+1$ equations

$$\begin{aligned} S_0^0 &= \rho(\omega_0)S_1^0(\omega_0)\mathbb{P}\{\omega_0\} + \dots + \rho(\omega_n)S_1^0(\omega_n)\mathbb{P}\{\omega_n\} = \mathbb{E}(\rho S_1^0) \\ \vdots &= \ddots \\ S_0^d &= \rho(\omega_0)S_1^d(\omega_0)\mathbb{P}\{\omega_0\} + \dots + \rho(\omega_n)S_1^d(\omega_n)\mathbb{P}\{\omega_n\} = \mathbb{E}(\rho S_1^d). \end{aligned}$$

Roughly speaking, one should expect to find a solution to these equations if there are at least as many unknowns as equations, i.e. if $n \geq d$. Otherwise, if $n < d$ then there are few unknowns than equations, and hence one should not expect to find a ρ satisfying the system.

Since the existence of the pricing kernel is equivalent to the lack of arbitrage, one has the following rule-of-thumb:

1FTAP:	No arbitrage	$\Leftrightarrow$	Existence of pricing kernel	"$\Leftrightarrow$	$n \geq d$"

Now, when is the market complete? Heuristically, we should expect there to be a unique pricing kernel ρ only if the number of unknowns equals the number of equations, i.e. $n = d$.

Or looking the other way, to replicate a contingent claim ξ_1 we need to find a $\bar{\pi}$ such that $\bar{\pi} \cdot \bar{S}_1 = \xi_1$. That is, we must solve the $n+1$ equations

$$\begin{aligned} \xi_1(\omega_0) &= \pi^0 S_1^0(\omega_0) + \dots + \pi^d S_1^d(\omega_0) \\ \vdots &= \ddots \\ \xi_1(\omega_n) &= \pi^0 S_1^0(\omega_n) + \dots + \pi^d S_1^d(\omega_n) \end{aligned}$$

in the $d+1$ unknowns $\pi^0, \dots, \pi^d$. There is usually exists a solution only if $n \leq d$.

Hence, we have the rule-of-thumb below:

2FTAP:	Market completeness	$\Leftrightarrow$	Uniqueness of pricing kernel	"$\Leftrightarrow$	$n = d$"

4. Pricing and hedging in an incomplete market by minimizing hedging error

In complete markets, we know how to compute the time-0 price of any contingent claim– just compute the expected value of the claim times the unique pricing kernel. Indeed, in a complete market, all contingent claims can be perfectly replicated, so their time-0 prices are just the time-0 prices of the corresponding replicating portfolio.

What does the no-arbitrage principle say about the prices of contingent claims in incomplete markets? Let $\bar{S}$ be a no-arbitrage market model.

Theorem (No-arbitrage price bounds). *There is no arbitrage in the augmented market* $(\bar{S}, \xi)$ *if*

$$\inf_\rho \mathbb{E}(\rho\xi_1) \leq \xi_0 \leq \sup_\rho \mathbb{E}(\rho\xi_1)$$

where the infimum and supremum are taken over pricing kernels ρ *such that* $\mathbb{E}(\rho|\xi_1|) < \infty$.

That is, the set of no-arbitrage prices is an interval, but unless the claim in attainable, the interval does not collapse to a single point. Then what to do? If you are the seller of a claim, what portfolio should you buy to minimize your exposure to the unhedgeable risk?

Of course, there are many, many answers to this question. One possible solution is to minimize expected square hedging error. Consider a market $\bar{S}$ and introduce a claim with payout ξ_1 at time 1. In this section, we suppose that the prices are square integrable $\mathbb{E}(|\bar{S}_1|^2) < \infty$ and $\mathbb{E}(\xi_1^2) < \infty$.

We are trying to find the attainable claim $X_1^* = \bar{\pi}^* \cdot \bar{S}_1$ which minimizes the functional $X_1 \mapsto \mathbb{E}[(\xi_1 - X_1)^2]$. That is, we find the wealth X_1^*, which can be attained by trading in the market, which is closest (in the least-squares sense) to the target claim ξ_1. Equivalently, we can think of X_1^* as the projection of ξ_1 on the subspace of attainable claims.

Hence if

$$X_1^* = \bar{\pi}^* \cdot \bar{S}_1$$

is optimal, we have the following interpretion:

- $\bar{\pi}^*$ the optimal vector of portfolio weights, and
- $\bar{\pi}^* \cdot \bar{S}_0 = X_0^*$ is the time-0 price of this portfolio.

In particular, the seller of the claim should charge at least X_0^* and invest the proceeds into the portfolio $\bar{\pi}^*$.

We can find the minimizer of the function

$$F(\bar{\pi}) = \mathbb{E}[(\xi_1 - \bar{\pi} \cdot \bar{S}_1)^2]$$

by the usual means of calculus:

$$\nabla F(\pi^*) = \mathbb{E}[\bar{S}_1(\xi_1 - \bar{\pi} \cdot \bar{S}_1)] = 0$$

yielding

$$\begin{aligned} \bar{\pi}* &= \mathbb{E}(V^{-1}\bar{S}_1\xi_1) \\ X_0^* &= \mathbb{E}[\bar{S}_0^{\mathrm{T}} V^{-1}\bar{S}_1\xi_1] \end{aligned}$$

and $V = \mathbb{E}[\bar{S}_1\bar{S}_1^{\mathrm{T}}]$ is a $(d+1) \times (d+1)$ matrix, assumed positive definite.

What can we say about this solution? Note that we can express the optimal initial wealth as

$$X_0^* = \mathbb{E}(\rho^*\xi_1)$$

where the random variable ρ^* is defined by

$$\rho^* = \bar{S}_0^{\mathrm{T}} V^{-1}\bar{S}_1$$

so that the pricing rule $\xi_1 \mapsto X_0^*$ is linear, just as in the case of a complete market.

The most important property of ρ^* is found in the following:

THEOREM. *The random variable $\rho^* = \bar{S}_0^{\mathrm{T}} V^{-1}\bar{S}_1$ minimizes the functional $\rho \mapsto \mathbb{E}(\rho^2)$ among all random variables ρ such that $\mathbb{E}(\rho\bar{S}_1) = \bar{S}_0$.*

PROOF. First we show that $\mathbb{E}(\bar{S}_1\rho^*) = \bar{S}_0$. We have the computation

$$\mathbb{E}(\bar{S}_1\rho^*) = \mathbb{E}[\bar{S}_1\bar{S}_1^{\mathrm{T}} V^{-1}\bar{S}_o] = \bar{S}_0.$$

It remains to show that ρ^* is minimal. Let ρ be another random variable such that $\mathbb{E}(\bar{S}_1\rho^*) = \bar{S}_0$. Then $\Delta\rho = \rho - \rho^*$ satisfies $\mathbb{E}(\bar{S}_1\Delta\rho) = 0$. Writing $\rho = \rho^* + \Delta\rho$, we have the computation

$$\begin{aligned}\mathbb{E}(\rho^2) &= \mathbb{E}[(\rho^* + \Delta\rho)^2] \\ &= \mathbb{E}[(\rho^*)^2] + 2\mathbb{E}[\bar{S}_0^{\mathrm{T}} V^{-1}\bar{S}_1\Delta\rho] + \mathbb{E}[(\Delta\rho)^2] \\ &\geq \mathbb{E}[(\rho^*)^2],\end{aligned}$$

completing the proof. □

Note that if ρ^* is a positive random variable, then ρ^* would be the pricing kernel with the smallest L^2 norm. In particular, if the market is complete, we have discovered an explicit formula for the unique pricing kernel.

However, for a general market $\rho*$ may well have the property $\mathbb{P}(\rho \leq 0) > 0$. In this case, ρ^* is not a pricing kernel. Indeed if ρ^* is not strictly positive and if one prices the claim by $\xi_0 = X_0^* = \mathbb{E}(\rho^*\xi_1)$, then it is possible that the augmented market has an arbitrage!

5. Change of numéraire and equivalent martingale measures

In this section, we introduce the important concepts of numéraire assets and equivalent martingale measures.

We need a definition:

Definition. Let $(\Omega, \mathcal{F})$ be a measurable space and let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $(\Omega, \mathcal{F})$. The measures $\mathbb{P}$ and $\mathbb{Q}$ are *equivalent*, written $\mathbb{P} \sim \mathbb{Q}$, if and only if

$$\mathbb{P}(A) = 1 \Leftrightarrow \mathbb{Q}(A) = 1$$

Example. Consider the sample space $\Omega = \{1, 2, 3\}$ with the set $\mathcal{F}$ of events all subsets of Ω. Consider probability measures $\mathbb{P}$ and $\mathbb{Q}$ defined by

- $\mathbb{P}\{1\} = 1/2, \mathbb{P}\{2\} = 1/2,$ and $\mathbb{P}\{3\} = 0$
- $\mathbb{Q}\{1\} = 1/1000, \mathbb{Q}\{2\} = 999/1000,$ and $\mathbb{Q}\{3\} = 0.$

Then $\mathbb{P}$ and $\mathbb{Q}$ are equivalent.

It turns out that equivalent measures can be characterized by the following theorem. When there are more than one probability measure floating around, we use the notation $\mathbb{E}^{\mathbb{P}}$ to denote expected value with respect to $\mathbb{P}$, etc.

Theorem (Radon–Nikodym theorem). *The probability measure $\mathbb{Q}$ is equivalent to the probability measure $\mathbb{P}$ if and only if there exists a positive random variable Z such that*

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}(Z\mathbb{1}_A)$$

for each $A \in \mathcal{F}$.

Note that $\mathbb{E}^{\mathbb{P}}(Z) = 1$ by putting $A = \Omega$ in the conclusion of theorem. Also, by the usual rules of integration theory, if ξ is a non-negative random variable then

$$\mathbb{E}^{\mathbb{Q}}(\xi) = \mathbb{E}^{\mathbb{P}}(Z\xi).$$

If $\mathbb{Q} \sim \mathbb{P}$, then the random variable Z is called the *density*, or the *Radon–Nikodym derivative*, of $\mathbb{Q}$ with respect to $\mathbb{P}$, and is often denoted

$$Z = \frac{d\mathbb{Q}}{d\mathbb{P}}.$$

In fact, $\mathbb{P}$ also has a density with respect to $\mathbb{Q}$ given by

$$\frac{d\mathbb{P}}{d\mathbb{Q}} = \frac{1}{Z}.$$

Now let's return to our financial model. To start off, we need the following definition:

DEFINITION. A *numéraire* is an asset with a strictly positive price at all times.

The idea is that a numéraire can be used to count money. Hence, we can speak in terms of prices relative to the numéraire. In particular, let ρ be a pricing kernel, and suppose asset i is a numéraire. Then, we have

$$\frac{S_0^j}{S_0^i} = \mathbb{E}\left[Z_i \frac{S_1^j}{S_1^i}\right]$$

where

$$Z_i = \rho \frac{S_1^i}{S_0^i}.$$

Notice that $Z_i > 0$ almost surely and $\mathbb{E}(Z_i) = 1$. Hence, we define an equivalent probability measure $\mathbb{Q}_i$ by

$$\mathbb{Q}_i(A) = \mathbb{E}(Z_i \mathbb{1}_A)$$

for all $A \in \mathcal{F}$. The measure $\mathbb{Q}_i$ is called an equivalent martingale measure:

DEFINITION. Let $\bar{S}$ be a market model defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The measure $\mathbb{P}$ is called the *objective* (or *historical* or *statistical*) *measure* for the model.

An *equivalent martingale measure* relative to the numéraire asset i is any probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ such that $\mathbb{E}^{\mathbb{Q}}(|S_1^j|/S_1^i|) < \infty$ and

$$\mathbb{E}^{\mathbb{Q}}\left(\frac{S_1^j}{S_1^i}\right) = \frac{S_0^j}{S_0^i}$$

for all $j \in \{0, \ldots, d\}$.

REMARK. As a preview of what's to come, the term equivalent martingale measure is appropriate since the stochastic processes $(S_t^j/S_t^i)_{t\in\{0,1\}}$ are a martingale for $\mathbb{Q}_i$ with respect to the filtration $(\mathcal{F}_t)_{t\in\{0,1\}}$ where $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_1 = \mathcal{F}$. We will elaborate on this in the multi-period case.

Unlike the notion of a pricing kernel, the notion of an equivalent martingale measure is 'numéraire-dependent'. That is, given a pricing kernel ρ we can construct an equivalent martingale measure corresponding to a different numéraire. That is, for every numéraire asset i, we can define a measure $\mathbb{Q}_i$ by

$$\frac{d\mathbb{Q}_i}{d\mathbb{P}} = \rho \frac{S_1^i}{S_0^i}.$$

In general, the measures $\mathbb{Q}_i$ constructed above are different for different i, even though they all correspond to the same pricing kernel ρ. The choice of numéraire is usually arbitrary, but there is one case that should be mentioned.

DEFINITION. A numéraire is *risk-free* if its time-1 price is not random. An equivalent martingale measure relative to a risk-free asset is called *risk-neutral.*

6. Discounted prices

From now on, we distinguish one asset and treat it differently than the d others. By convention, we choose asset 0, but this choice is arbitrary. To make the notation easier, we let

$$B_t = S_t^0, \text{ and } S_t = (S_1^1, \ldots, S_t^d)$$

from now on. As mentioned earlier, asset 0 is often cash, in which case $B_0 = B_1 = 1$ or a bank (or money market) account in which case $B_0 = 1$ and $B_1 = 1 + r$, where r is the interest rate. The actual identity of asset 0 is irrelevant for much of our analysis, as long as it is a numéraire. This brings us to a standing assumption:

ASSUMPTION. Asset 0 is a numéraire, i.e. $B_0 > 0$ and $B_1 > 0$ almost surely.

The remaining d assets are often called *risky assets*, their riskiness being relative to the numéraire asset 0.

If the investor's initial wealth X_0 is fixed, the budget constraint means that we cannot freely choose the investor's portfolio. We write

$$\phi = \pi^0, \text{ and } \pi = (\pi^1, \ldots, \pi^d),$$

so that

$$\begin{aligned} X_0 &= \phi B_0 + \pi \cdot S_0, \\ X_1 &= \phi B_1 + \pi \cdot S_1. \end{aligned}$$

Since we have assumed that asset 0 is a numéraire, we can solve for ϕ, yielding

$$\phi = \frac{X_0}{B_0} - \pi \cdot \frac{S_0}{B_0}.$$

Plugging this into the self-financing condition yields after some manipulation

$$X_1 = \frac{B_1}{B_0} X_0 + \pi \cdot \left(S_1 - \frac{B_1}{B_0} S_0 \right)$$

To clean things up, we count money in units of asset 0: Define the new quantities for both $t \in \{0, 1\}$

$$\tilde{X}_t = \frac{X_t}{B_t}, \text{ and } \tilde{S}_t = \frac{S_t}{B_t}$$

denoting the wealth and the risky asset prices *discounted* by the numéraire asset 0. (Of course, if asset 0 were cash then $B_0 = B_1 = 1$ almost surely, and we wouldn't need the new notation.)

In this new notation, the budget constraint and self-financing condition neatly combine to yield

$$\boxed{\tilde{X}_1 = \tilde{X}_0 + \pi \cdot (\tilde{S}_1 - \tilde{S}_0)}.$$

Now it is time to consider the notion of arbitrage in this new notation.

Definition. An *arbitrage* (relative to the numéraire asset 0) is a portfolio $\pi \in \mathbb{R}^d$ of risky assets such that

$$\mathbb{P}[\pi \cdot (\tilde{S}_1 - \tilde{S}_0) \geq 0] = 1$$
$$\mathbb{P}[\pi \cdot (\tilde{S}_1 - \tilde{S}_0) > 0] > 0$$

Remark. Suppose $\pi \in \mathbb{R}^d$ is an arbitrage relative to asset 0. Then $\bar{\pi} = (\phi, \pi) \in \mathbb{R}^{d+1}$ is an arbitrage by our old definition of the word, once we let $\phi = -\pi \cdot \tilde{S}_0$. Indeed, in this case, the initial wealth is $X_0 = \phi B_0 + \pi \cdot S_0 = 0$ and $X_1 = B_1 \pi \cdot (\tilde{S}_1 - \tilde{S}_0)$, so $X_1 \geq 0$ a.s. and $\mathbb{P}(X_1 > 0) > 0$.

Conversely, if the portfolio $\bar{\pi} = (\phi, \pi)$ is an arbitrage according to our old definition, then $\tilde{X}_1 \geq 0 \geq \tilde{X}_0$ a.s. and $\mathbb{P}(\tilde{X}_1 > \tilde{X}_0) > 0$. But $\tilde{X}_1 - \tilde{X}_0 = \pi \cdot (\tilde{S}_1 - \tilde{S}_0)$, so π is an arbitrage relative to asset 0.

Of course, we are actually interested in the lack of arbitrage:

the market model has no arbitrage	if and only if $\pi \cdot (\tilde{S}_1 - \tilde{S}_0) \geq 0$ a.s. implies $\pi \cdot (\tilde{S}_1 - \tilde{S}_0) = 0$.

We can now rewrite the first fundamental theorem:

Theorem (First Fundamental Theorem of Asset Pricing). *The market model has no arbitrage if and only if there exists an equivalent martingale measure (relative to asset 0).*

Note that an equivalent martingale measure $\mathbb{Q}$ is simply a probability measure $\mathbb{Q} \sim \mathbb{P}$ such that

$$\mathbb{E}^{\mathbb{Q}}(\tilde{S}_1) = \tilde{S}_0.$$

We can also redefine the term attainable:

Definition. A claim with payout ξ_1 is *attainable* if there is a real number x and portfolio $\pi \in \mathbb{R}^d$ of risky assets such that

$$\tilde{\xi}_1 = x + \pi \cdot (\tilde{S}_1 - \tilde{S}_0)$$

where $\tilde{\xi}_1 = \xi_1 / B_1$.

Attainable claims can be characterized in terms of equivalent martingale measures:

Theorem (Characterization of attainable claims). *A claim with payout ξ_1 is attainable if and only if there exists a constant x such that*

$$\mathbb{E}^{\mathbb{Q}}(\tilde{\xi}_1) = x$$

for all equivalent martingale measure $\mathbb{Q}$ such that $\mathbb{E}^{\mathbb{Q}}(|\tilde{\xi}_1|) < \infty$, where $\tilde{\xi}_1 = \xi_1 / B_1$.

Finally, the second fundamental theorem can be rewritten:

Theorem (Second Fundamental Theorem of Asset Pricing). *A market model is complete if and only if the equivalent martingale measure (relative to asset 0) is unique.*

These are the formulations that we will use for the remainder of these notes.

CHAPTER 2

Multi-period models

Now that the financial foundation has been laid in the one-period models, we can proceed briskly into the natural generalization of multi-period discrete time. Essentially, this entails replacing the time index set $\{0, 1\}$ with the non-negative integers $\mathbb{Z}_+$ and keeping track of the resulting complications.

1. The set-up

So we consider a market with $d + 1$ assets. The prices of these assets are modelled by the stochastic process $(B, S) = (B_t, S_t)_{t \in \mathbb{Z}_+}$ defined on some background probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Our first goal is to find the proper generalization to this setting of the assumption that the time-0 prices B_0 and S_0 are constant.

We need to formalize the concept of information being revealed as time marches forward. The correct notions are that of measurablility and of a filtration.

Example. Consider the experiment of tossing a coin two times. We can model this experiment on the sample space $\Omega = \{HH, HT, TH, TT\}$. Pick an outcome $\omega \in \Omega$ at random.

At time 0, the coin has not been tossed, so there are only two questions you can answer.

> Is ω in $\emptyset$? No.
> Is ω in Ω? Yes.

but

> Is ω in $\{HH, HT\}$? I can assign a probability $\mathbb{P}\{HH, HT\} = 1/2$, but I can't answer the question with certainty.

At time 1, the coin has been tossed once. Then you can always answer these four questions.

> Is ω in $\emptyset$? No.
> Is ω in Ω? Yes.
> Is ω in $\{HH, HT\}$? Yes, if the first toss is heads, no otherwise.
> Is ω in $\{TH, TT\}$? Yes, if the first toss is tails, no otherwise.

but

> Is ω in $\{HH, HT, TH\}$? Yes, if the first toss is heads, but what if the first toss comes up tails? So, *in general* I can't answer the question with certainty after observing the first flip.

After the second toss, of course, you can answer every question. So, the flow of information is modelled by the following sigma-fields

- $\mathcal{F}_0 = \{\emptyset, \Omega\}$,

- $\mathcal{F}_1 = \{\emptyset, \{HH, HT\}, \{TH, TT\}, \Omega\}$,
- $\mathcal{F}_2 =$ the set of all sixteen subsets of Ω.

Now consider a stochastic process $(Y_t)_{t\in\{0,1,2\}}$ that has the property that the value of the random variable Y_t is known once after t tosses of the coin.

For instance, Y_0 must be a constant,

$$Y_0(\omega) = a \text{ for all } \omega \in \Omega,$$

since there is no information before the experiment. On the other hand, the random variable Y_1 must be of the form

$$Y_1(\omega) = \begin{cases} a & \text{if } \omega \in \{HH, HT\} \\ b & \text{if } \omega \in \{TH, TT\} \end{cases}$$

since the only information known at time 1 is whether or not the first coin came up heads. Finally, Y_2 can be any function on Ω, that is, of the form

$$Y_2(\omega) = \begin{cases} a & \text{if } \omega = HH \\ b & \text{if } \omega = HT \\ c & \text{if } \omega = TH \\ d & \text{if } \omega = TT. \end{cases}$$

Notice that for all $t \in \{0, 1, 2\}$ the event $\{Y_t \le x\}$ is in $\mathcal{F}_t$ for every $x \in \mathbb{R}$.

With this motivation, we have the following definitions:

Definition. Let $\mathcal{G} \subseteq \mathcal{F}$ be a sigma-field. A random variable $\xi : \Omega \to \mathbb{R}$ is *measurable* with respect to $\mathcal{G}$ (or briefly, $\mathcal{G}$-measurable) if and only if the event $\{\xi \le x\}$ is an element of $\mathcal{G}$ for all $x \in \mathbb{R}$. If ξ is a random vector, then $\xi = (\xi_1, \ldots, \xi_n)$ is $\mathcal{G}$-measurable if and only if ξ_i is $\mathcal{G}$-measurable for each $i \in \{1, \ldots, n\}$.

Let $\mathbb{T} \subseteq \mathbb{R}_+$ be an index set. For this chapter, we have $\mathbb{T} = \mathbb{Z}_+$ the positive integers, but the some of the following definitions are stated in more generality than we need here to avoid repetition in later chapters in which $\mathbb{T} = \mathbb{R}_+$.

Definition. A *filtration* $(\mathcal{F}_t)_{t\in\mathbb{T}}$ is an increasing collection of sigma-fields on Ω such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ if $s \le t$.

Remark. When dealing with filtrations, we will nearly always assume that $\mathcal{F}_0$ is trivial:

$$A \in \mathcal{F}_0 \Leftrightarrow \mathbb{P}(A) = 0 \text{ or } \mathbb{P}(A) = 1.$$

In particular, all $\mathcal{F}_0$ measurable random variables are almost surely constant.

Definition. A stochastic process $(Y_t)_{t\in\mathbb{T}}$ is *adapted* to a filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$ iff Y_t is $\mathcal{F}_t$-measurable for each $t \in \mathbb{T}$.

Now we can state the proper generalization of the assumption in one period models that the time-0 asset prices are constant: We equip the background probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}_+}$. The sigma-field $\mathcal{F}_t$ is our model of what information is known to the market participants at time t.

Assumption. The stochastic process $(B, S) = (B_t, S_t)_{t\in\mathbb{Z}_+}$ is assumed to be adapted to $(\mathcal{F}_t)_{t\in\mathbb{Z}_+}$.

Remark. Since the only random variables measurable with respect to the trivial sigma-field are constants, we see that B_0 and S_0 are constants as before.

Our second assumption is natural, generalizes the assumption made in one period models, and doesn't really need further comment.

Assumption. Asset 0 is a numéraire so that $B_t > 0$ almost surely for all $t \in \mathbb{Z}_+$.

As before, given this market model, we introduce an investor. Now there are multiple periods, so we need to be careful about when an investor chooses his investment portfolio.

- X_t denotes the investor's wealth at the beginning of period t.
- ϕ_t denotes the number of shares of asset 0 held between periods $t-1$ and t.
- π_t denotes the portfolio of risky assets held between periods $t-1$ and t.

The investor's wealth and porfolio are connected by the following relationships:

$$\boxed{\begin{array}{rcll} X_{t-1} & = & \phi_t B_{t-1} + \pi_t \cdot S_{t-1} & \text{the budget constraint} \\ X_t & = & \phi_t B_t + \pi_t \cdot S_t & \text{the self-financing condition} \end{array}}$$

Note that we should assume that our investor is not clairvoyant. Hence we henceforth only consider portfolios (ϕ_t, π_t) that are $\mathcal{F}_{t-1}$-measurable. Such processes have a name:

Definition. A stochastic process $(Y_t)_{t\in\mathbb{N}}$ is *predictable* if Y_t is $\mathcal{F}_{t-1}$-measurable for each $t \in \mathbb{N}$.

Remark. Note that the time index set for a predictable process $(Y_t)_{t\in\mathbb{N}}$ is (usually) $\mathbb{N} = \{1, 2, \ldots\}$, not $\mathbb{Z}_+$. Hence Y_0 is not necessarily defined.

Just as we did in the one-period case, we can solve for ϕ_t:

$$\phi_t = \frac{X_{t-1}}{B_{t-1}} - \pi_t \cdot \frac{S_{t-1}}{B_{t-1}}.$$

Now count money in units of the numéraire asset 0:

$$\tilde{X}_t = \frac{X_t}{B_t} \quad \text{and} \quad \tilde{S}_t = \frac{S_t}{B_t}$$

so that the budget constraint and self-financing condition combine to yield

$$\tilde{X}_t = \tilde{X}_{t-1} + \pi_t \cdot (\tilde{S}_t - \tilde{S}_{t-1}).$$

The final formula to remember is then

$$\boxed{\tilde{X}_t = \tilde{X}_0 + \sum_{s=1}^{t} \pi_s \cdot (\tilde{S}_s - \tilde{S}_{s-1}).}$$

Remark. For a fixed $t \in \mathbb{N}$ we refer to the d-dimensional random vector π_t as the investor's portfolio; the d-dimensional stochastic process $(\pi_t)_{t\in\mathbb{N}}$ is called the investor's *trading strategy*.

2. The first fundamental theorem of asset pricing

Following the discussion from before we can define an arbitrage.

DEFINITION. An *arbitrage* is a strategy $(\pi_t)_{t\in\mathbb{N}}$ with the property that there exists a (non-random) time $T \in \mathbb{N}$ such that

$$\begin{aligned}\mathbb{P}\left(\sum_{s=1}^{T} \pi_s \cdot (\tilde{S}_s - \tilde{S}_{s-1}) \geq 0\right) &= 1 \\ \mathbb{P}\left(\sum_{s=1}^{T} \pi_s \cdot (\tilde{S}_s - \tilde{S}_{s-1}) > 0\right) &> 0.\end{aligned}$$

Before we can state the first fundamental theorem, we have to recall some results and definitions about conditional expectations and martingale theory.

THEOREM (Existence and uniqueness of conditional expectations). *Let X be a integrable random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-sigma-field of $\mathcal{F}$. Then there exists an integrable $\mathcal{G}$-measurable random variable Y such that*

$$\mathbb{E}(\mathbb{1}_G Y) = \mathbb{E}(\mathbb{1}_G X)$$

for all $G \in \mathcal{G}$. Furthermore, if there exists another $\mathcal{G}$-measurable random variable Y' such that $\mathbb{E}(\mathbb{1}_G Y') = \mathbb{E}(\mathbb{1}_G X)$ for all $G \in \mathcal{G}$, then $Y = Y'$ almost surely.

DEFINITION. Let X be an integrable random variable and let $\mathcal{G} \subset \mathcal{F}$ be a sigma-field. The *conditional expectation* of X given $\mathcal{G}$, written $\mathbb{E}(X|\mathcal{G})$, is a $\mathcal{G}$-measurable random variable with the property that

$$\mathbb{E}\left[\mathbb{1}_G \mathbb{E}(X|\mathcal{G})\right] = \mathbb{E}(\mathbb{1}_G X)$$

for all $G \in \mathcal{G}$.

EXAMPLE. (Sigma-field generated by a countable partition) Let X be a non-negative random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $G_1, G_2, \ldots$ be a sequence of disjoint events with $\mathbb{P}(G_n) > 0$ for all $n \in \mathbb{N}$ and $\bigcup_{n\in\mathbb{N}} G_n = \Omega$.

Let $\mathcal{G}$ be the smallest sigma-field containing $\{G_1, G_2, \ldots, ...\}$. Then

$$\mathbb{E}(X|\mathcal{G})(\omega) = \frac{\mathbb{E}(X\mathbb{1}_{G_n})}{\mathbb{P}(G_n)} \text{ if } \omega \in G_n.$$

This example relates the notion of conditional expection given a sigma-field and that of conditional expectation given an event, since we usually write $\frac{\mathbb{E}(X\mathbb{1}_G)}{\mathbb{P}(G)} = \mathbb{E}(X|G)$.

More concretely, suppose $\Omega = \{HH, HT, TH, TT\}$ consists of two tosses of a coin, and let $\mathcal{G} = \{\emptyset, \{HH, HT\}, \{TH, TT\}, \Omega\}$ be the sigma-field containg the information revealed by the first toss. Suppose the coin is fair, so that each outcome is equally likely. Consider the random variable

$$\xi(\omega) = \begin{cases} a & \text{if } \omega = HH \\ b & \text{if } \omega = HT \\ c & \text{if } \omega = TH \\ d & \text{if } \omega = TT. \end{cases}$$

Then

$$\mathbb{E}(\xi|\mathcal{G})(\omega) = \begin{cases} (a+b)/2 & \text{if } \omega \in \{HH, HT\} \\ (c+d)/2 & \text{if } \omega \in \{TH, TT\} \end{cases}$$

The important properties of conditional expectations are collected below:

THEOREM. *Let all random variables appearing below be such that the relevant conditional expectations are defined, and let $\mathcal{G}$ be a sub-sigma-field of the sigma-field $\mathcal{F}$ of all events.*

- *linearity:* $\mathbb{E}(aX + bY|\mathcal{G}) = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$ *for all constants a and b*
- *positivity: If $X \geq 0$ almost surely, then $\mathbb{E}(X|\mathcal{G}) \geq 0$ almost surely, with almost sure equality if and only if $X = 0$ almost surely.*
- *Jensen's inequality: If f is convex, then* $\mathbb{E}[f(X)|\mathcal{G}] \geq f[\mathbb{E}(X|\mathcal{G})]$
- *monotone convergence theorem: If $0 \leq X_n \uparrow X$ a.s. then $\mathbb{E}(X_n|\mathcal{G}) \uparrow \mathbb{E}(X|\mathcal{G})$ a.s.*
- *Fatou's lemma: If $X_n \geq 0$ a.s. for all n, then* $\mathbb{E}(\liminf_n X_n|\mathcal{G}) \leq \liminf_n \mathbb{E}(X_n|\mathcal{G})$
- *dominated convergence theorem: If $\sup_n |X_n|$ is integrable and $X_n \to X$ a.s. then $\mathbb{E}(X_n|\mathcal{G}) \to \mathbb{E}(X|\mathcal{G})$ a.s.*
- *If X is independent of $\mathcal{G}$ (the events $\{X \leq x\}$ and G are independent for each $x \in \mathbb{R}$ and $G \in \mathcal{G}$) then $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$. In particular, $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$ if $\mathcal{G}$ is trivial.*
- *'take-out-what's-known': If X is $\mathcal{G}$-measurable, then $\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G})$. In particular, if X is $\mathcal{G}$-measurable, then $\mathbb{E}(X|\mathcal{G}) = X$.*
- *tower property or law of iterated expectations: If $\mathcal{H} \subseteq \mathcal{G}$ then*

$$\mathbb{E}[\mathbb{E}(X|\mathcal{G})|\mathcal{H}] = \mathbb{E}[\mathbb{E}(X|\mathcal{H})|\mathcal{G}] = \mathbb{E}(X|\mathcal{H})$$

As hinted at in Chapter 1, the most important concept in financial mathematics is that of a martingale. A martingale is simply an adapted stochastic process that is constant on average in a following sense:

DEFINITION. A *martingale* relative to a filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$ is a stochastic process $M = (M_t)_{t\in\mathbb{T}}$ with the following properties:

- $\mathbb{E}(|M_t|) < \infty$ for all $t \in \mathbb{T}$
- $\mathbb{E}(M_t|\mathcal{F}_s) = M_s$ for all $0 \leq s \leq t$.

REMARK. If $\mathbb{T} = \mathbb{Z}_+$, it is an exercise to show that an integrable process M is a martingale only if $\mathbb{E}(M_{t+1}|\mathcal{F}_t) = M_t$ for all $t \geq 0$. That is, it is sufficient to verify the conditional expectations of the process one period ahead.

Below are some examples of martingales. Before listing them, it is convenient to introduce a definition:

DEFINITION. Given a stochastic process $Y = (Y_t)_{t\in\mathbb{T}}$, let $\mathcal{F}_t$ be the smallest sigma-field for which the random variables Y_s is measurable for all $0 \leq s \leq t$. The *natural filtration* of Y is the filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$, that is, the smallest filtration for which ξ is adapted.

In what follows, if a stochastic process is given but a filtration is not explicitly mentioned, then we are implicitly working with the natural filtration of the process.

EXAMPLE. Let $X_1, X_2, X_3, \ldots$ be independent integrable random variables such that $\mathbb{E}(X_i) = 0$ for all $i \in \mathbb{N}$. The process $(S_t)_{t\in\mathbb{Z}_+}$ given by $S_0 = 0$ and

$$S_t = X_1 + \ldots + X_t$$

is a martingale relative to its natural filtration. Indeed, the random variable S_t is integrable since

$$\mathbb{E}(|S_t|) \leq \mathbb{E}(|X_1|) + \ldots + \mathbb{E}(|X_t|)$$

by the triangular inequality. Also,

$$\begin{aligned}\mathbb{E}(S_{t+1}|\mathcal{F}_t) &= \mathbb{E}(S_t + X_{t+1}|\mathcal{F}_n)\\ &= \mathbb{E}(S_t|\mathcal{F}_t) + \mathbb{E}(X_{t+1}|\mathcal{F}_t)\\ &= S_t + \mathbb{E}(X_{t+1}) = S_t.\end{aligned}$$

The following theorem shows how to take one martingale and build another one.

THEOREM. *Let M be a martingale and let H be a bounded predictable process. Then the process N defined by*

$$N_t = \sum_{s=1}^{t} H_s(M_s - M_{s-1})$$

is a martingale.

PROOF. By assumption, we have $\mathbb{E}(|M_t|) < \infty$ and there exist a constant $C > 0$ such that $|H_t| < C$ almost surely for all $t \in \mathbb{Z}_+$. Hence

$$\begin{aligned}\mathbb{E}(|N_t|) &\leq \sum_{s=1}^{t} \mathbb{E}(|H_s||M_s - M_{s-1}|)\\ &\leq \sum_{s=1}^{t} C[\mathbb{E}(|M_s|) + \mathbb{E}(|M_{s-1}|)] < \infty\end{aligned}$$

Since

$$\mathbb{E}(N_{t+1} - N_t|\mathcal{F}_t) = \mathbb{E}(H_{t+1}(M_{t+1} - M_t)|\mathcal{F}_t) = H_{t+1}\mathbb{E}(M_{t+1} - M_t|\mathcal{F}_t) = 0$$

we're done. □

REMARK. The process N in the theorem is often called a *martingale transform* or a *discrete time stochastic integral*, .

EXAMPLE. We now construct one of the most important examples of a martingale. Let X be an integrable random variable, and let

$$M_t = \mathbb{E}(X|\mathcal{F}_t).$$

Then $M = (M_t)_{t\in\mathbb{T}}$ is a martingale. First,

$$\begin{aligned}\mathbb{E}(|M_t|) &= \mathbb{E}\{|\mathbb{E}(X|\mathcal{F}_t)|\}\\ &\leq \mathbb{E}\{\mathbb{E}(|X||\mathcal{F}_t)\}\\ &= \mathbb{E}(|X|) < \infty\end{aligned}$$

Now,

$$\begin{aligned}\mathbb{E}(M_t|\mathcal{F}_s) &= \mathbb{E}[\mathbb{E}(X|\mathcal{F}_t)|\mathcal{F}_s] \\ &= \mathbb{E}(X|\mathcal{F}_s) = M_s\end{aligned}$$

by the tower property.

We now are ready to consider our model $\bar{S} = (B_t, S_t)_{t\in\{0,\dots,T\}}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and adapted to a filtration $(\mathcal{F}_t)_{t\in\{0,\dots,T\}}$. Recall the notation $\tilde{S}_t = S_t/B_t$ for the discounted risky asset prices.

Definition. An *equivalent martingale measure* is a measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ such that $\tilde{S}$ is a martingale for $\mathbb{Q}$.

Theorem (First Fundamental Theorem of Asset Pricing). *The market model has no arbitrage if and only if there exists an equivalent martingale measure.*

Proof. We only consider the easier direction, that the existence of an equivalent martingale measure implies the lack of arbitrage, in the case where $T = 2$. The idea is the same as in the $T = 1$ case, but one has to be a bit more careful. Let π be such that $\tilde{X}_t = \sum_{s=1}^t \pi_s \cdot \Delta\tilde{S}_s$ satifies $\tilde{X}_2 \geq 0$ a.s. Our aim is to show $\tilde{X}_2 = 0$ a.s.

Let $\mathbb{Q}$ be an equivalent martingale measure. Note $\mathbb{Q}(\tilde{X}_2 \geq 0) = 1$ because $\mathbb{P} \sim \mathbb{Q}$. We would be done if we could show $\mathbb{E}^{\mathbb{Q}}(\tilde{X}_2) = 0$, because this would imply that $\mathbb{Q}(\tilde{X}_2 = 0) = 1$ and hence $\mathbb{P}(\tilde{X}_2 = 0) = 1$, again by equivalence of measures.

To do this, let $A_n = \{|\tilde{X}_1| \leq n, |\pi_2| \leq n\}$. Note A_n is in $\mathcal{F}_1$. Furthermore, note that

$$\mathbb{1}_{A_n}\tilde{X}_2 = \mathbb{1}_{A_n}\tilde{X}_1 + \mathbb{1}_{A_n}\pi_2 \cdot \Delta\tilde{S}_2$$

is $\mathbb{Q}$-integrable, since the first term is bounded, and the second term is bounded by $n|\Delta\tilde{S}_2|$, which is $\mathbb{Q}$-integrable since $\tilde{S}$ is a $\mathbb{Q}$-martingale. Now,

$$\begin{aligned}0 \leq \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{A_n}\tilde{X}_2|\mathcal{F}_1] &= \mathbb{1}_{A_n}\tilde{X}_1 + \mathbb{1}_{A_n}\pi_2 \cdot \mathbb{E}(\Delta\tilde{S}_2|\mathcal{F}_1) \\ &= \mathbb{1}_{A_n}\tilde{X}_1\end{aligned}$$

and letting $n \to \infty$, we have $\tilde{X}_1 \geq 0$ a.s. Now we can compute

$$\mathbb{E}^{\mathbb{Q}}(\tilde{X}_1) = \pi_1 \cdot \mathbb{E}(\tilde{S}_1) = 0$$

to conclude $\tilde{X}_1 = 0$ a.s. Finally, since

$$\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{A_n}\tilde{X}_2|\mathcal{F}_1] = \mathbb{1}_{A_n}\tilde{X}_1 = 0$$

we conclude $\mathbb{1}_{A_n}\tilde{X}_2 = 0$ a.s. for each n. Letting $n \to \infty$ shows $\tilde{X}_2 = 0$ a.s. as desired. □

3. European and American contingent claims

Now given our multi-period market model, we can introduce a contingent claim. Unlike the one-period case, there are now essentially two types of contingent claims: those, called *European* that mature at a fixed date and those, called *American* contingent claims, that can be *exercised* whenever the holder of the claim wants.

3.1. European claims. We concentrate on the European claims first. The following theory should seem very familiar since there really is nothing new.

Let $(B, S) = (B_t, S_t)_{t\in\mathbb{Z}_+}$ be an arbitrage-free market. We now introduce a (European) contingent claim which matures at time $T \in \mathbb{N}$. We model the payout of the claim by an $\mathcal{F}_T$-measurable random variable ξ_T. Think of the example of a call option on a stock maturing at time T with strike K. In this case $\xi_T = (S_T - K)^+$.

THEOREM. *The augmented market $(B_t, S_t, \xi_t)_{t\in\{0,\dots,T\}}$ is free of arbitrage if and only if there exists an equivalent martingale measure $\mathbb{Q}$ such that*

$$\xi_t = \mathbb{E}^{\mathbb{Q}}\left(\frac{B_t}{B_T}\xi\big|\mathcal{F}_t\right)$$

for all $t \in \{0, \dots, T\}$.

PROOF. By the first fundamental theorem of asset pricing, there is no-arbitrage if and only if there exists an equivalent measure $\mathbb{Q}$ such that $(\tilde{S}_t, \tilde{\xi}_t)_{t\in\{0,\dots,T\}}$ is a martingale for $\mathbb{Q}$, where $\tilde{\xi}_t = \xi_t/B_t$. But $(\tilde{\xi}_t)_{t\in\{0,\dots,T\}}$ is a martingale if and only if

$$\tilde{\xi}_t = \mathbb{E}^{\mathbb{Q}}(\tilde{\xi}_T|\mathcal{F}_t)$$

and we're done. □

The above theorem gives a lot of flexibility in pricing contingent claims unless the claim is attainable, or the set $\mathcal{Q}$ of equivalent martingale measures contains just one element.

DEFINITION. A claim maturing at time $T \in \mathbb{N}$ is *attainable* if and only if its payout ξ_T is an $\mathcal{F}_T$-measurable random variable such that

$$\tilde{\xi}_T = x + \sum_{s=1}^{T} \pi_s \cdot (\tilde{S}_s - \tilde{S}_{s-1})$$

for some $x \in \mathbb{R}$ and predictable strategy $(\pi_t)_{t\in\mathbb{N}}$, where $\tilde{\xi}_T = \frac{\xi_t}{B_T}$.

The market is *complete* if and only if every claim is attainable.

The following should now come as no surprise.

THEOREM (Characterization of attainable claims). *Given an arbitrage free market model $\bar{S}$, a contingent claim with payout ξ_T is attainable if and only if there a constant x such that*

$$x = \mathbb{E}^{\mathbb{Q}}(\tilde{\xi}_T)$$

for all equivalent martingale measures $\mathbb{Q}$ such that $\mathbb{E}^{\mathbb{Q}}(|\tilde{\xi}_T|) < \infty$.

PROOF. We just consider the case $T = 2$, and suppose that the claim is attainable, so that there exists a constant x and strategy $(\pi)_{t=\{1,2\}}$ such that $X_2 = \xi_2$ a.s. where

$$\tilde{X}_t = x + \sum_{s=1}^{t} \pi_s \cdot (\tilde{S}_s - \tilde{S}_{s-1})$$

We need to show $x = \mathbb{E}^{\mathbb{Q}}(\tilde{\xi}_2)$ for all $\mathbb{Q}$ such that $\tilde{\xi}_2$ is integrable. (It needs to be proven that there is at least one $\mathbb{Q}$ with this property, but we do not do so here.)

As before, let $A_n = \{|\tilde{X}_1| \leq n, |\pi_2| \leq n\}$ and hence

$$\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{A_n}\tilde{X}_2|\mathcal{F}_1] = \mathbb{1}_{A_n}\tilde{X}_1.$$

Now, since $|\mathbb{1}_{A_n}\tilde{X}_2| \leq \tilde{X}_2$ and this is integrable by assumption, we can use the conditional dominated convergence theorem on the left side to let $n \to \infty$ to get

$$\mathbb{E}^{\mathbb{Q}}[\tilde{X}_2|\mathcal{F}_1] = \tilde{X}_1.$$

Now $\tilde{X}_1 = x + \pi_1(\tilde{S}_1 - \tilde{S}_0)$ is integrable and has mean x, since π_1 is a constant. Applying the tower law yields

$$\mathbb{E}^{\mathbb{Q}}[\mathbb{E}^{\mathbb{Q}}(\tilde{X}_2|\mathcal{F}_1)] = \mathbb{E}(\tilde{X}_1) = x.$$

□

Finally, we state the characterization of complete markets.

THEOREM (Second Fundamental Theorem of Asset Pricing). *The market is complete if and only if there exists a unique equivalent martingale measure.*

3.2. American claims. We now discuss American claims. Here, things are quite different.

The canonical example of an American claim is the American put option– a contract which gives the buyer the right (but not the obligation) to sell the underlying stock at a fixed strike price $K > 0$ at *any time* between time 0 and a fixed maturity date $T \in \mathbb{N}$. Hence, the payout of the option is $(K - S_\tau)^+$ where $\tau \in \{0, \dots, T\}$ is a time chosen by the holder of the put to exercise the option.

Hence, the payout of an American claim is specified by two ingredients:

- a maturity date $T \in \mathbb{N}$,
- an adapted process $(\xi_t)_{t\in\{0,\dots,T\}}$.

For instance, in the case of an American put, we may take $\xi_t = (K - S_t)^+$. Unlike the European claim, the holder of an American claim can choose to exercise the option at any time τ before or at maturity. However, to rule out clairvoyance, we insist that τ is a stopping time:

DEFINITION. A *stopping time* for a filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$ is a random variable τ taking values in $\mathbb{T} \cup \{\infty\}$ such that the event $\{\tau \leq t\}$ is $\mathcal{F}_t$-measurable for all $t \in \mathbb{T}$.

EXAMPLE. (First passage time) Here is a typical example of a stopping time. Let $(Y_t)_{t\in\mathbb{Z}_+}$ be an adapted process and fix $a \in \mathbb{R}$. Then the random variable

$$\tau = \inf\{t \in \mathbb{Z}_+ : Y_t > a\}$$

corresponding to the first time the process crosses the level a is a stopping time. Indeed,

$$\{\tau > t\} = \{Y_s \leq a \text{ for all } s = 0, \dots, t\}$$

is $\mathcal{F}_t$-measurable, and hence so is $\{\tau > t\}^c = \{\tau \leq t\}$.

Now, if an American claim matures at $T \in \mathbb{N}$ and is specified by the payout process $(\xi_t)_{t\in\{0,\dots,T\}}$, then the actual payout of the claim is modelled by the random variable ξ_τ, where τ is any stopping time for the filtration taking values in $\{0, \dots, T\}$.

We can think of the American claim then as a family, indexed by the stopping time τ, of European claims with payouts ξ_τ. To simplify matters, we make the following assumption in this subsection:

$$\text{The market model } (B,S) = (B_t, S_t)_{t\in\mathbb{Z}_+} \text{ is complete.}$$

The unique equivalent martingale measure is denoted $\mathbb{Q}$.

Intuitively, the seller of such a claim should at time 0 charge at least the amount

$$\sup_{\tau\in\mathcal{T}} \mathbb{E}^{\mathbb{Q}}\left(\frac{B_0}{B_\tau}\xi_\tau\right)$$

to be sure that he can hedge the option, where the supremum is taken over the set $\mathcal{T}$ of stopping times smaller than or equal to T and where $\mathbb{Q}$ is the unique martingale measure. Indeed, this is the case.

Theorem. *Consider a complete market (B,S), and suppose that the adapted process $(\xi_t)_{t\in\{0,\dots,T\}}$ specifies the payout of an American claim maturing at $T \in \mathbb{N}$.*

There exists a trading strategy $(\pi_t)_{t\in\{1,\dots,t\}}$ with corresponding discounted wealth process

$$\tilde{X}_t = \tilde{X}_0 + \sum_{s=1}^{t} \pi_s \cdot (\tilde{S}_s - \tilde{S}_{s-1})$$

such that

$$X_t \geq \xi_t$$

almost surely for all $t \in \{0,\dots,T\}$ if and only if

$$X_0 \geq \sup_{\tau\in\mathcal{T}} \mathbb{E}^{\mathbb{Q}}\left(\frac{B_0}{B_\tau}\xi_\tau\right).$$

Remark. The theorem says that if the initial wealth is sufficiently large, the investor can *super-replicate* the payout of the American claim.

The rest of this subsection is dedicated to proving this theorem.

We need some new vocabulary:

Definition. A *supermartingale* relative to a filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$ is an adapted stochastic process $(U_t)_{t\in\mathbb{T}}$ with the following properties:

- $\mathbb{E}(|U_t|) < \infty$ for all $t \in \mathbb{T}$
- $\mathbb{E}(U_t|\mathcal{F}_s) \leq U_s$ for all $0 \leq s \leq t$.

A *submartingale* is an adapted process $(V_t)_{t\in\mathbb{Z}_+}$ with the following properties:

- $\mathbb{E}(|V_t|) < \infty$ for all $t \in \mathbb{T}$
- $\mathbb{E}(V_t|\mathcal{F}_s) \geq V_s$ for all $t \in \mathbb{T}$.

Remark. Hence a supermartingale *decreases* on average, while a submartingale *increases* on average. A martingale is a stochastic process that is both a supermartingale and a submartingale.

Now we need a result of general interest:

THEOREM (Doob decomposition theorem). *Let U be a supermartingale indexed by $\mathbb{Z}_+$. Then there is a unique decomposition*

$$U_t = M_t - A_t$$

where M is a martingale and A is a predictable non-decreasing process with $A_0 = 0$.

PROOF. Let $M_0 = U_0$ and define

$$M_{t+1} - M_t = U_{t+1} - \mathbb{E}(U_{t+1}|\mathcal{F}_t) \text{ for } t \in \mathbb{N}.$$

Then M is a martingale. Now let $A_0 = 0$ and

$$A_{t+1} - A_t = U_t - \mathbb{E}(U_{t+1}|\mathcal{F}_t) \text{ for } t \in \mathbb{N}.$$

Clearly A_{t+1} is $\mathcal{F}_t$-measurable and the process A is non-decreasing process since U is a supermartingale. Summing up,

$$\begin{aligned} M_t - A_t &= M_0 - A_0 + \sum_{s=1}^{t}(M_s - M_{s-1} - A_s + A_{s-1}) \\ &= U_0 + \sum_{s=1}^{t}(U_s - U_{s-1}) \\ &= U_t. \end{aligned}$$

To show uniqueness, assume that $U_t = M_t - A_t = M'_t - A'_t$. Then $M - M'$ is a predictable martingale, that is, a constant. □

DEFINITION. Let $(Y_t)_{t\in\{0,\dots,T\}}$ be a given integrable adapted process. Define an adapted process $(U_t)_{t\in\{0,\dots,T\}}$ by

$$\begin{aligned} U_T &= Y_T \\ U_t &= \max\{Y_t, \mathbb{E}(U_{t+1}|\mathcal{F}_t)\} \text{ for } t \in \{0,\dots,T-1\} \end{aligned}$$

The process $(U_t)_{t\in\{0,\dots,T\}}$ is called the *Snell envelope* of $(Y_t)_{t\in\{0,\dots,T\}}$.

REMARK. The Snell envelope clearly satisfies both

$$U_t \geq Y_t \text{ and } U_t \geq \mathbb{E}(U_{t+1}|\mathcal{F}_t)$$

almost surely. Thus, another way to describe the Snell envelope of a process is to say it is the smallest supermartingale dominating that process.

In our application $(Y_t)_{t\in\{0,\dots,T\}}$ will be the process specifying the discounted payout of the given American claim $Y_t = \tilde{\xi}_t$.

THEOREM. *Let $(Y_t)_{t\in\{0,\dots,T\}}$ be an adapted process, let $(U_t)_{t\in\{0,\dots,T\}}$ be its Snell envelope with Doob decomposition $U_t = M_t - A_t$. Let*

$$\tau^* = \min\{t \in \{0,\dots,T\} : A_{t+1} > 0\}$$

with the convention $\tau^* = T$ on $\{A_t = 0$ for all $t\}$. Then τ^* is a stopping time and $Y_{\tau^*} = M_{\tau^*}$.

Proof. That τ^* is a stopping time follows from the fact that the non-decreasing process $(A_t)_{t\in\{0,\dots,T\}}$ is predictable.

Now note that

$$\mathbb{E}(U_{t+1}|\mathcal{F}_t) = \mathbb{E}(M_{t+1} - A_{t+1}|\mathcal{F}_t) = M_t - A_{t+1}$$

so that by the definition of Snell envelope

$$M_t - A_t = \max\{Y_t, M_t - A_{t+1}\}.$$

In particular,

$$M_{\tau^*} = \max\{Y_{\tau^*}, M_{\tau^*} - A_{\tau^*+1}\}$$

since $A_{\tau^*} = 0$. But since $A_{\tau^*+1} > 0$ we must conclude

$$U_{\tau^*} = M_{\tau^*} = Y_{\tau^*}.$$

□

Theorem. *Let Y be an adapted process indexed by $\{0,\dots,T\}$, and let U be its Snell envelope. Then*

$$U_0 = \sup_{\tau\in\mathcal{T}} \mathbb{E}(Y_\tau).$$

Proof. Since U is a supermartingale,

$$U_0 \geq \mathbb{E}(U_\tau)$$

by the optional sampling theorem. (See example sheet 2.) But letting $\tau^* = \min\{t \in \{0,\dots,T\} : A_{t+1} > 0\}$ where $U = M - A$ is the Doob decomposition of U, we have

$$U_0 = M_0 = \mathbb{E}(M_{\tau^*}) = \mathbb{E}(Y_{\tau^*}).$$

□

Definition. A stopping time τ such that $\mathbb{E}(Y_\tau) = U_0$ is called an *optimal stopping time.*

.

Returning to finance, let $(\xi_t)_{t\in\{0,\dots,T\}}$ be the process specifying the payout of an American option, and let $(U_t)_{t\in\{0,\dots,T\}}$ be the Snell envelope of the discounted payout $\tilde{\xi}$ with Doob decomposition $U_t = M_t - A_t$.

We now will use the assumption that the market is complete: consider an investor with initial discounted wealth

$$x = U_0 = M_0 = \sup_{\tau\in\mathcal{T}} \mathbb{E}^{\mathbb{Q}}(\tilde{\xi}_\tau).$$

By the complete market assumption, we can find a predictable trading strategy $(\pi_t)_{t\in\{1,\dots,T\}}$ so that

$$x + \sum_{s=1}^{T} \pi_s \cdot (\tilde{S}_s - \tilde{S}_{s-1}) = M_T.$$

Since $(\tilde{S}_t)_{t\in\{0,\dots,T\}}$ is a martingale for $\mathbb{Q}$, [and since $(\pi_t)_{t\in\{1,\dots,T\}}$ is bounded because we're in a complete market and hence any $\mathcal{F}_t$-measurable random variable can only take a finite number of values] the investor's discounted wealth $\tilde{X}$ is a martingale, where

$$\tilde{X}_t = x + \sum_{s=1}^{t} \pi_s \cdot (\tilde{S}_s - \tilde{S}_{s-1}).$$

Hence $\tilde{X}_t = M_t$ for all $t \in \{0, \dots, T\}$ and the trading strategy $(\pi_t)_{t\in\{1,\dots,t\}}$ super-replicates the payout of the American claim, since

$$\tilde{\xi}_t \le U_t \le M_t = \tilde{X}_t$$

almost surely for all $t \in \{0, \dots, T\}$.

We have shown that if the seller of the claim has $x = U_0 = M_0$ intial discounted wealth, he can super-replicate the payout of the American claim. Furthermore, he cannot do so with a smaller endowment, as there exists a stopping time τ^* such that $X_{\tau^*} = \xi_{\tau^*}$.

4. Locally equivalent measures

In all the examples in this course, we will deal with finite horizon problems. However, it is sometimes more convenient not to explicitly mention the horizon. In particular, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space equipped with a filtration $(\mathcal{F}_t)_{t\ge\mathbb{T}}$. In this section, we will briefly discuss a weakening of the notion of equivalence of measures that is suitable for our purposes.

To see where we're going, suppose that the measure $\mathbb{Q}$ is equivalent to $\mathbb{P}$. By the Radon–Nikodym theorem, there exists a positive random variable $Z = \frac{d\mathbb{Q}}{d\mathbb{P}}$ such that $\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}(Z\mathbb{1}_A)$. We can associate with any equivalent measure a positive martingale given by

$$Z_t = \mathbb{E}^{\mathbb{P}}(Z|\mathcal{F}_t).$$

Notice that if $A \in \mathcal{F}_t$ then

$$\begin{aligned}
\mathbb{Q}(A) &= \mathbb{E}^{\mathbb{P}}(\mathbb{1}_A Z) \\
&= \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}(Z\mathbb{1}_A|\mathcal{F}_t)] \\
&= \mathbb{E}^{\mathbb{P}}[\mathbb{1}_A\mathbb{E}^{\mathbb{P}}(Z|\mathcal{F}_t)] \\
&= \mathbb{E}^{\mathbb{P}}[\mathbb{1}_A Z_t].
\end{aligned}$$

What the above calculation shows is that the density of the restriction $\mathbb{Q}|_{\mathcal{F}_t}$ of the measure $\mathbb{P}$ to the sub-sigma-field $\mathcal{F}_t \subseteq \mathcal{F}$ with respect to $\mathbb{P}|_{\mathcal{F}_t}$ is the random variable Z_t.

Now we are in a position to generalize the notion of equivalent measures.

Definition. The measure $\mathbb{Q}$ is *locally* equivalent to $\mathbb{P}$ if and only if the restriction $\mathbb{Q}|_{\mathcal{F}_t}$ of the measure $\mathbb{Q}$ to the sigma-field $\mathcal{F}_t$ is equivalent to $\mathbb{P}|_{\mathcal{F}_t}$ for each $t \in \mathbb{T}$.

Theorem (Radon–Nikodym, local version). *The measures $\mathbb{P}$ and $\mathbb{Q}$ are locally equivalent if and only if there exists a positive martingale Z with $Z_0 = 1$ such that*

$$\frac{d\mathbb{Q}|_{\mathcal{F}_t}}{d\mathbb{P}|_{\mathcal{F}_t}} = Z_t.$$

We will call the martingale Z the *density process* for $\mathbb{Q}$ with respect to $\mathbb{P}$.

To conclude this chapter, let us consider a financial model $(B, S) = (B, S)_{t\in\mathbb{Z}_+}$. If no final time horizon $T > 0$ is explicitly mentioned, then we will extend the definition of equivalent martingale measure to include measures $\mathbb{Q}$ locally equivalent to $\mathbb{P}$ such that $\tilde{S} = S/B$ is a $\mathbb{Q}$-martingale.

CHAPTER 3

Brownian motion and stochastic calculus

In the lectures up to now, we have considered investors whose discounted wealth process is typically of the form

$$\tilde{X}_t = \tilde{X}_0 + \sum_{s=1}^{t} \pi_s \cdot (\tilde{S}_s - \tilde{S}_{s-1}).$$

In this chapter, we consider the limit as trading frequency becomes more and more frequent, and hence we need to understand the continuous time generalization

$$\tilde{X}_t = \tilde{X}_0 + \int_0^t \pi_s \cdot d\tilde{S}_s.$$

We will see that the above *stochastic integral* can, in fact, be defined.

Now recall that the first fundamental theorem of asset pricing tells us, in the discrete-time world, that if there is no-arbitrage, then there is an equivalent measure $\mathbb{Q}$ such that $(\tilde{S}_t)_{t\in\mathbb{Z}_+}$ is a martingale. As a preview of what's to come, we will see that essentially all continuous martingales $M = (M_t)_{t\in\mathbb{R}_+}$ are of the form

$$M_t = M_0 + \int_0^t \alpha_s \cdot dW_s$$

where $W = (W_t)_{t\in\mathbb{R}_+}$ is a Brownian motion. If the sample paths of Brownian motion were differentiable, we would be able to define the stochastic integral $\int_0^t \alpha_s \cdot dW_s$ by $\int_0^t \alpha_s \cdot \frac{dW_s}{ds} ds$ and the story would be over. Unfortunately, the sample paths of Brownian motion are not differentiable, and so we will have to do more work to make sense of the situation. So although the stochastic integral behaves in some ways like a Riemann–Stieltjes integral of calculus, a word of warning is in order:

> The rules of stochastic calculus are not the same as those of ordinary calculus.

Our goals will be to define a Brownian motion, to construct the Brownian stochastic integral, and to learn the rules of the resulting calculus. The following chapter will provide an extremely brief introduction to this theory.

1. Brownian motion

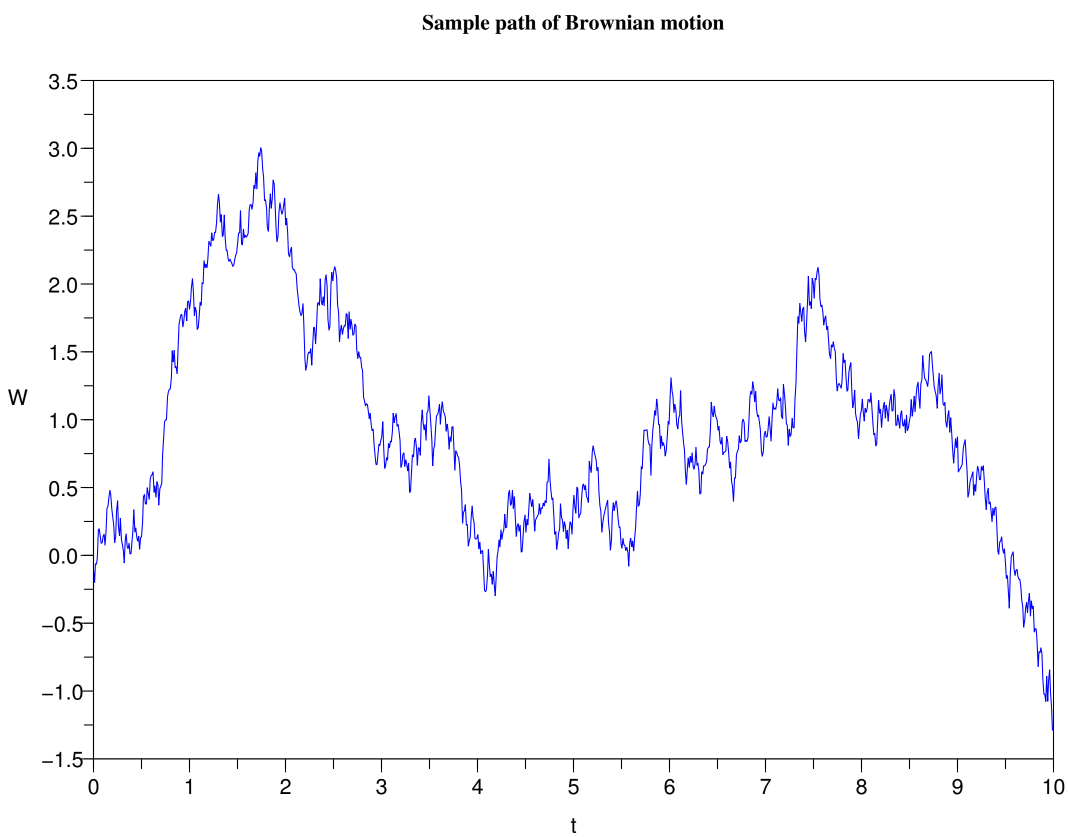
In this section, we introduce one of the most fundamental continuous-time stochastic processes, Brownian motion. As hinted above, our primary interest in this process is that it will be the building block for all of the continuous-time market models studied in these lectures.

Definition. A *Brownian motion* $W = (W_t)_{t\in\mathbb{R}_+}$ is a collection of random variables such that

- $W_0(\omega) = 0$ for all $\omega \in \Omega$,
- for all $0 \le t_0 < t_1 < \ldots < t_n$ the increments $W_{t_{i+1}} - W_{t_i}$ are independent, and the distribution of $W_t - W_s$ is $N(0, |t-s|)$,
- the sample path $t \mapsto W_t(\omega)$ is continuous all $\omega \in \Omega$.

It is not clear that Brownian motion exists. That is, does there exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which the uncountable collection of random variables $(W_t)_{t\in\mathbb{R}_+}$ can be simultaneously defined in such a way that the above definition holds? The answer, of course, is yes, and the proof of this fact is due to Wiener in 1930. Therefore, the Brownian motion is also often called the *Wiener process*, especially in the U.S.

Although the sample paths of Brownian motion are continuous, they are very irregular. Below is a computer simulation of a one-dimensional Brownian motion:



The following is a very important result, due to Lévy, which can quantify the irregularity of the Brownian sample path. In this chapter, we will use the phrase 'a sequence of partitions of $[0, t]$ with vanishing norm' to mean a collection of points $0 = t_0^{(N)} \le t_1^{(N)} \ldots \le t_N^{(N)} = t$

such that $\max_{n\in\{1,\dots,N\}}|t_n^{(N)}-t_{n-1}^{(N)}|\to 0$ as $N\to\infty$. Useful properties of such sequences is

$$\sum_{n=1}^N (t_n^{(N)}-t_{n-1}^{(N)}) = t$$

and

$$\sum_{n=1}^N (t_n^{(N)}-t_{n-1}^{(N)})^2 \le \max_{n\in\{1,\dots,N\}}|t_n^{(N)}-t_{n-1}^{(N)}|\sum_{n=1}^N (t_n^{(N)}-t_{n-1}^{(N)}) \to 0.$$

THEOREM. *Let W and $W^\perp$ be independent one-dimensional Brownian motions. For every sequence of partitions of $[0,t]$ with vanishing norm we have*

$$\sum_{n=1}^N (W_{t_n^{(N)}}-W_{t_{n-1}^{(N)}})^2 \to t \quad \textit{in probability}$$

and

$$\sum_{n=1}^N (W_{t_n^{(N)}}-W_{t_{n-1}^{(N)}})(W^\perp_{t_n^{(N)}}-W^\perp_{t_{n-1}^{(N)}}) \to 0 \quad \textit{in probability.}$$

REMARK. For comparison, consider a continuously differentiable function $f:\mathbb{R}_+\to\mathbb{R}$. Then, we have

$$\begin{aligned}\sum_{n=1}^N [f(t_n^{(N)})-f(t_{n-1}^{(N)})]^2 &= \sum_{n=1}^N f'(s_n^{(N)})^2(t_n^{(N)}-t_{n-1}^{(N)})^2 \\ &\le \max_{s\in[0,t]} f'(s)^2 \sum_{n=1}^N (t_n^{(N)}-t_{n-1}^{(N)})^2 \to 0\end{aligned}$$

where $t_{n-1}^{(N)}<s_n^{(N)}<t_n^{(N)}$ by the mean value theorem. We can conclude that from the above theorem that a typical Brownian sample path is not a continuously differentiable function of time.

PROOF. By definition, the increments of Brownian motion are Gaussian randoms so that

$$\mathbb{E}[(W_t-W_s)^2] = t-s$$

and

$$\text{Var}[(W_t-W_s)^2] = 2(t-s)^2$$

for every $0\le s\le t$. Hence

$$\mathbb{E}\left[\sum_{n=1}^N (W_{t_n^{(N)}}-W_{t_{n-1}^{(N)}})^2\right] = \sum_{n=1}^N (t_n^{(N)}-t_{n-1}^{(N)}) = t$$

and, by the independence of the increments of Brownian motion,

$$\text{Var}\left[\sum_{n=1}^N (W_{t_n^{(N)}}-W_{t_{n-1}^{(N)}})^2\right] = 2\sum_{n=1}^N (t_n^{(N)}-t_{n-1}^{(N)})^2 \to 0$$

and the first conclusion follows from Chebychev's inequality.

Since

$$\mathbb{E}[(W_t - W_s)(W_t^\perp - W_s^\perp)] = 0$$

and

$$\mathrm{Var}[(W_t - W_s)(W_t^\perp - W_s^\perp)] = (t-s)^2$$

the second conclusion follows similarly. □

The Brownian motion is too rough to apply the Riemann–Stieltjes integration theory. Fortunately, there is an integration theory that does the job. It is based on the fact that Brownian motion is a martingale.

THEOREM. *Let W be a scalar Brownian motion, and let $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ be its natural filtration. Then W is a martingale.*

PROOF. Testing integrability is easy since Gaussian random variables have finite exponential moments. Now fix $0 \leq s < t$.

$$\begin{aligned} \mathbb{E}(W_t|\mathcal{F}_s) &= \mathbb{E}(W_s + W_t - W_s|\mathcal{F}_s) \\ &= W_s + \mathbb{E}(W_t - W_s) \\ &= W_s \end{aligned}$$

Note that the fact that $W_t - W_s$ is independent of $\mathcal{F}_s$ is used to pass from a conditional expectation to an unconditional expectation. □

2. Itô stochastic integration

We now have sufficient motivation to construct a stochastic integral with respect to a Wiener process. What follows is the briefest of sketches of the theory. There are now plenty of places to turn for a proper treatment of the subject. For instance, please consult one of the following references:

- L.C.G. Rogers and D. Williams, *Diffusions, Markov Processes, and Martingales: Volume 2*
- I. Karatzas and S.E. Shreve, *Brownian Motion and Stochastic Calculus.*

2.1. The L^2 theory. To get things started, let W be a scalar Brownian motion. We will assume that W is adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$. Of this filtration we will assume that it satisfies what are called the *usual conditions* of right-continuity $\mathcal{F}_t = \bigcap_{\epsilon>0} \mathcal{F}_{t+\epsilon}$ and that $\mathcal{F}_0$ contains all $\mathbb{P}$-null events. We also will assume that for each $0 \leq s < t$ the increment $W_t - W_s$ is independent of $\mathcal{F}_s$.

The first building block of the theory are the simple predictable integrands.

DEFINITION. A *simple predictable integrand* is an adapted process $\alpha = (\alpha_t)_{t\in\mathbb{R}_+}$ of the form

$$\alpha_t(\omega) = \sum_{n=1}^{N} \mathbb{1}_{(t_{n-1},t_n]}(t) a_n(\omega)$$

where a_n is bounded and $\mathcal{F}_{t_{n-1}}$-measurable for some $0 \leq t_0 < t_1 < ... < t_N < \infty$. For simple predictable integrands we define the stochastic integral by the formula

$$\int_0^\infty \alpha_s dW_s = \sum_{n=1}^{N} a_n (W_{t_n} - W_{t_{n-1}})$$

Theorem (Itô's isometry). *For a simple predictable integrand α, we have*

$$\mathbb{E}\left[\left(\int_0^\infty \alpha_s dW_s\right)^2\right] = \mathbb{E}\left(\int_0^\infty \alpha_s^2 ds\right)$$

Proof.

$$\mathbb{E}\left[\left(\int_0^\infty \alpha_s dW_s\right)^2\right] = \mathbb{E}\sum_{m,n} a_m a_n (W_{t_m} - W_{t_{m-1}})(W_{t_n} - W_{t_{n-1}})$$

Consider the terms in the sum when $m \leq n$. Applying the tower property, we see

$$\begin{aligned}\mathbb{E}a_m a_n (W_{t_m} - W_{t_{m-1}})(W_{t_n} - W_{t_{n-1}}) &= \mathbb{E}[\mathbb{E}(a_m a_n (W_{t_m} - W_{t_{m-1}})(W_{t_n} - W_{t_{n-1}})|\mathcal{F}_{t_{n-1}})] \\ &= \mathbb{E}[a_m a_n \mathbb{E}((W_{t_m} - W_{t_{m-1}})(W_{t_n} - W_{t_{n-1}})|\mathcal{F}_{t_{n-1}})]\end{aligned}$$

since a_m and a_n are $\mathcal{F}_{t_{n-1}}$-measurable.

If $m < n$,

$$\mathbb{E}[(W_{t_m} - W_{t_{m-1}})(W_{t_n} - W_{t_{n-1}})|\mathcal{F}_{t_{n-1}}] = (W_{t_m} - W_{t_{m-1}})\mathbb{E}[W_{t_n} - W_{t_{n-1}}] = 0$$

and if $m = n$

$$\mathbb{E}[(W_{t_n} - W_{t_{n-1}})^2|\mathcal{F}_{t_{n-1}}] = \mathbb{E}[(W_{t_n} - W_{t_{n-1}})^2] = t_n - t_{n-1}$$

since the increment $W_{t_n} - W_{t_{n-1}}$ is independent of $\mathcal{F}_{t_{n-1}}$. In particular, the off diagonal terms cancel, and we have

$$\begin{aligned}\mathbb{E}\sum_{m,n} a_m a_n (W_{t_m} - W_{t_{m-1}})(W_{t_n} - W_{t_{n-1}}) &= \mathbb{E}\sum_n a_n^n (t_n - t_{n-1}) \\ &= \mathbb{E}\left(\int_0^\infty \alpha_s^2 ds\right)\end{aligned}$$

as desired. □

Now, the map defined by $I(\alpha) = \int_0^\infty \alpha_s dW_s$ is an isometry from the space of simple predictable integrands to the space L^2 of square-integrable random variables. The fact that L^2 is complete is the key observation which allows us to build the stochastic integral of more general integrands.

Definition. The *predictable sigma-field* $\mathcal{P}$ is the sigma-field on the product space $\mathbb{R}_+ \times \Omega$ generated by sets of the form $(s,t] \times A$ where $0 \leq s < t$ and A is $\mathcal{F}_s$-measurable. Equivalently, the predictable sigma-field is that generated by the simple, predictable integrands.

A *predictable* process α is a map $\alpha : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ that is $\mathcal{P}$-measurable. Equivalently, predictable processes are limits of simple, predictable integrands.

Remark. Every left-continuous, adapted process is predictable. These are the examples to keep in mind, since they are the ones that come up most in application.

Now, suppose $(\alpha^{(k)})_{k\in\mathbb{N}}$ is a sequence of simple predictable integrands converging to a predictable process α in the sense that

$$\mathbb{E}\left(\int_0^\infty (\alpha_s^{(k)} - \alpha_s)^2 ds\right) \to 0$$

as $k \to \infty$. By Itô's isometry the sequence $I(\alpha^{(k)})$ is a Cauchy sequence, which by the completeness of L^2, converges to some random variable. This is what we take as the definition.

DEFINITION. If α is predictable and

$$\mathbb{E}\left(\int_0^\infty \alpha_s^2 ds\right) < \infty$$

then

$$\int_0^\infty \alpha_s dW_s = L^2 - \lim_k \int_0^\infty \alpha_s^{(k)} dW_s$$

where $\alpha^{(k)}$ is any sequence of simple, predictable integrands converging in L^2 to α.

But please note:

> The stochastic integral is *not* defined as the *almost sure* limit of a sequence of Riemann–Stieltjes integrals!!

Of course, we are not really interested in integrals over the whole interval $[0, \infty)$ but rather finite intervals $[0, t]$. This is easily handled.

DEFINITION.

$$\int_0^t \alpha_s dW_s = \int_0^\infty \alpha_s \mathbb{1}_{\{0 < s \leq t\}} dW_s$$

whenever the right-hand side is well-defined.

What can we say, within the L^2-theory, about a process defined by a stochastic integral?

THEOREM. *Let*

$$M_t = \int_0^t \alpha_s dW_s$$

where α is predictable and

$$\mathbb{E}\left(\int_0^t \alpha_s^2 \, ds\right) < \infty$$

for each $t \geq 0$. Then M is a continuous martingale.

2.2. Localization. In this section, we show how to extend the definition of stochastic integral to predictable processes α such that

$$\int_0^t \alpha_s^2 \, ds < \infty \text{ almost surely}$$

for all $t \geq 0$. The technique is called *localization.*

Define the stopping times

$$\tau_n = \inf\left\{t \geq 0 : \int_0^t \alpha_s^2 \, ds = n\right\}$$

for each $n \in \mathbb{N}$, where $\inf \emptyset = \infty$ as usual, and let

$$\alpha_t^{(n)} = \alpha_t \mathbb{1}_{\{t \leq \tau_n\}}.$$

Note that since $\mathbb{E}\left(\int_0^t (\alpha_s^{(n)})^2 ds\right) \leq n$, the process $\left(\int_0^t \alpha_s^{(n)} dW_s\right)_{t\in\mathbb{R}_+}$ is well-defined by the L^2 theory.

Now fix $t > 0$ and define the increasing sequence of events $A_n = \{\omega \in \Omega : \tau_n \geq t\}$. Since $\int_0^t \alpha_s^2 \, ds < \infty$ almost surely for all $t \geq 0$, we have $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = 1$. Hence we can define the stochastic integral by the formula

$$\int_0^t \alpha_s dW_s = \lim_{n\to\infty} \int_0^t \alpha_s^{(n)} dW_s$$

where the limit is in probability.

The process $L = \left(\int_0^t \alpha_s dW_s\right)_{t\in\mathbb{R}_+}$ defined in this way is still continuous, but in general there is no guarantee that L a martingale. Indeed, the stochastic integral defined by the localized version of the L^2 stochastic integration theory may not be a martingale, but by construction it is what is called a local martingale:

Definition. An adapted process $(L_t)_{t\in\mathbb{R}_+}$ is called a *local martingale* if and only if there exists an increasing sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ with $\tau_n \to \infty$ almost surely such that the stopped process $(L_{t\wedge\tau_n})_{t\in\mathbb{R}_+}$ is a martingale for all $n \in \mathbb{N}$.

To summarize:

If α is an adapated, left-continuous process such that $\int_0^t \alpha_s^2 \, ds < \infty$ almost surely for each $t \geq 0$ then $M_t = \int_0^t \alpha_s dW_s$ is defined. In all cases, M is a continuous local martingale. If in addition we have $\mathbb{E}\left(\int_0^t \alpha_s^2 \, ds\right) < \infty$ then M is a *true* martingale (as opposed to a being strictly local martingale).

Remark. If you're ambitious, you will now try to build a stochastic integration theory starting with a general continuous local martingale L, rather than Brownian motion. The steps are the same. First you localize L, so you can assume L is a square-integrable martingale. Then, you can do the L^2 theory as before, but Itô's isometry becomes

$$\mathbb{E}\left[\left(\int_0^\infty \alpha_s dL_s\right)^2\right] = \mathbb{E}\left[\int_0^\infty \alpha_s^2 d\langle L\rangle_s\right]$$

where

$$\langle L\rangle_t = \lim_{N\to\infty} \sum_{n=1}^{N} (L_{t_n^{(N)}} - L_{t_{n-1}^{(N)}})^2.$$

Then you can define the integral over finite intervals, and finally, extend the integrands by localization. Once you're done, you will have built a stochastic integration theory that has the very nice property that the integral $I_t = \int_0^t \alpha_s dL_s$ is defined when $\int_0^t \alpha_s^2 d\langle L\rangle_s, \infty$ a.s., and proces I is another continuous local martingale.

3. Itô's formula

In the last section, we sketched the constructed of a stochastic integral with respect to a Wiener process. What makes the Itô stochastic integral useful is that there is a corresponding stochastic calculus. The basic building block of this calculus is the chain rule, called Itô's formula.

3.1. The scalar version. Let $(W_t)_{t\in\mathbb{R}_+}$ be a scalar Brownian motion adapted to a filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying our usual conditions.

Let $(\alpha_t)_{t\in\mathbb{R}_+}$ and $(k_t)_{t\in\mathbb{R}_+}$ be predictable real-valued processes such that

$$\int_0^t \alpha_s^2 ds < \infty \text{ and } \int_0^t |k_s| ds < \infty$$

almost surely for all $t \geq 0$. Fix a real number X_0 and let

$$X_t = X_0 + \int_0^t \alpha_s dW_s + \int_0^t k_s ds.$$

Both integrals appearing the above equation now be interpreted, the first as a stochastic integral and the second as a pathwise Lebesgue integral. A process $(X_t)_{t\in\mathbb{R}_+}$ of the above form is often called an *Itô process.*

We are now ready for the first version of Itô's formula:

Theorem (Itô's formula, scalar version). *Let $f : \mathbb{R} \to \mathbb{R}$ be twice continuously differentiable. Then*

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)\alpha_s dW_s + \int_0^t (f'(X_s)k_s + \frac{1}{2} f''(X_s)\alpha_s^2) ds.$$

Let us highlight a difference between Itô and ordinary calculus, by noting the mysterious appearance of the f'' term in Itô's formula. This term would not appear in the chain rule of ordinary calculus. But consider the example $f(x) = x^2$ so that

$$W_t^2 = 2\int_0^t W_s dW_s + t.$$

Note that since $\mathbb{E}\left(\int_0^t W_s^2 ds\right) = t^2/2 < \infty$, the stochastic integral $\left(\int_0^t W_s dW_s\right)_{t\in\mathbb{R}_+}$ is a *true* martingale. Can you verify, directly from the definition of Brownian motion, that the process $(W_t^2 - t)_{t\in\mathbb{R}_+}$ is a martingale?

We now introduce some notions which helps with computations involving Itô's formula.

Definition. A *semimartingale* is a process X of the form

$$X_t = A_t + L_t$$

where L is a local martingale and A is a process of bounded variation. (For instance, an Itô process is a semimartingale.)

For a semimartingale X, the *quadratic variation* process $\langle X \rangle = (\langle X \rangle_t)_{t\in\mathbb{R}_+}$, if it exists, is defined by the limit in probability

$$\langle X \rangle_t = \lim_{N\to\infty} \sum_{n=1}^N (X_{t_n^{(N)}} - X_{t_{n-1}^{(N)}})^2$$

over a sequence of partitions with vanishing norm. If $(Y_t)_{t\in\mathbb{R}_+}$ is another semimartingale, then the *quadratic covariation* process $\langle X, Y \rangle = (\langle X, Y \rangle_t)_{t\in\mathbb{R}_+}$ is defined by the limit

$$\langle X, Y \rangle_t = \lim_{N\to\infty} \sum_{n=1}^N (X_{t_n^{(N)}} - X_{t_{n-1}^{(N)}})(Y_{t_n^{(N)}} - Y_{t_{n-1}^{(N)}}).$$

THEOREM. *The map $(X,Y) \mapsto \langle X,Y\rangle$ is bilinear, and the following table summarizes its possible values, where $(a_t)_{t\in\mathbb{R}_+}$ and $(b_t)_{t\in\mathbb{R}_+}$ are processes such that the relevant integrals are defined, and where $(W_t)_{t\in\mathbb{R}_+}$ and $(W_t^\perp)_{t\in\mathbb{R}_+}$ are independent Brownian motions.*

$\left\langle \int_0^\cdot a_s\, ds, \int_0^\cdot b_s\, ds \right\rangle_t = 0$	$\left\langle \int_0^\cdot a_s\, ds, \int_0^\cdot b_s\, dW_s \right\rangle_t = 0$
$\left\langle \int_0^\cdot a_s\, dW_s, \int_0^\cdot b_s\, dW_s \right\rangle_t = \int_0^t a_s b_s\, ds$	$\left\langle \int_0^\cdot a_s\, dW_s, \int_0^\cdot b_s\, dW_s^\perp \right\rangle_t = 0$

EXAMPLE. For instance, suppose we have two Itô processes

$$\begin{aligned} X_t &= X_0 + \int_0^t \alpha_s^{(1)}\, dW_s^{(1)} + \int_0^t \alpha_s^{(2)}\, dW_s^{(2)} + \int_0^t h_s\, ds \\ Y_t &= Y_0 + \int_0^t \beta_s^{(1)}\, dW_s^{(1)} + \int_0^t \beta_s^{(2)}\, dW_s^{(2)} + \int_0^t k_s\, ds \end{aligned}$$

for independent Brownian motions $W^{(1)}$ and $W^{(2)}$, then the quadratic covariation can be computed by bilinearity:

$$\langle X,Y\rangle_t = \int_0^t (\alpha_s^{(1)}\beta_s^{(1)} + \alpha_s^{(2)}\beta_s^{(2)})\, ds.$$

Armed with this the notion of quadratic variation, we are ready to see an indication for why Itô's formula may be true. Let

$$dX_t = \alpha_t\, dW_t + k_t\, dt$$

where the differential notation is shorthand for the corresponding integrals. Notice that in this differential notation, symbol $d\langle X\rangle_t$ means

$$d\langle X\rangle_t = \alpha_t^2\, dt.$$

Fix a partion of $[0,t]$ and consider the following second order Taylor approximation:

$$\begin{aligned} f(X_t) - f(X_0) &= \sum_{n=1}^N f(X_{t_n}) - f(X_{t_{n-1}}) \\ &\approx \sum_{n=1}^N f'(X_{t_{n-1}})(X_{t_n} - X_{t_{n-1}}) + \frac{1}{2} f''(X_{t_{n-1}})(X_{t_n} - X_{t_{n-1}})^2 \\ &\approx \int_0^t f'(X_s) dX_s + \int_0^t \frac{1}{2} f''(X_s) d\langle X\rangle_s \end{aligned}$$

In fact, it is customary to write out Itô's formula as

$$\boxed{df(X_t) = f'(X_t)dX_t + \frac{1}{2} f''(X_t) d\langle X\rangle_t}$$

where the differential notation should really be intrepreted as the corresponding integral equation.

Example. Consider the Itô process given by

$$X_t = X_0 + aW_t + bt$$

for some constants $a, b \in \mathbb{R}$. Letting

$$Y_t = e^{X_t},$$

we would like to show that the process $(Y_t)_{t\in\mathbb{R}_+}$ is an Itô process, and write down its decomposition in terms of ordinary and stochastic integrals.

Let $f(x) = e^x$. Then $f'(x) = e^x$ and $f''(x) = e^x$. Also,

$$dX_t = a\ dW_t + b\ dt \quad \text{and} \quad d\langle X\rangle_t = a^2\ dt$$

So Itô's formula says:

$$\begin{aligned} df(X_t) &= f'(X_t)dX_t + \frac{1}{2}f''(X_t)d\langle X\rangle_t \\ \Rightarrow dY_t &= Y_t[(b + a^2/2)dt + a\ dW_t] \end{aligned}$$

Example. Suppose we have the Itô process given implicitly as the solution of the stochastic differential equation

$$dY_t = Y_t[\beta dt + \alpha\ dW_t]$$

for some constants $\alpha, \beta \in \mathbb{R}$. Letting

$$Z_t = \log(Y_t),$$

we know from the previous example that $Z_t = Z_0 + aW_t + bt$, where $a = \alpha$ and $b = \beta - \alpha^2/2$. But to be safe, let's check. Let $g(y) = \log(y)$. Then $g'(y) = \frac{1}{y}$ and $g''(y) = -\frac{1}{y^2}$, and also

$$d\langle Y\rangle_t = \alpha^2 Y_t^2\ dt.$$

Again, Itô's formula says:

$$\begin{aligned} dg(Y_t) &= g'(Y_t)\ dY_t + \frac{1}{2}g''(Y_t)\ d\langle Y\rangle_t \\ \Rightarrow dZ_t &= \frac{1}{Y_t}Y_t[\beta dt + \alpha\ dW_t] + \frac{1}{2}\left(-\frac{1}{Y_t^2}\right)\alpha^2 Y_t^2\ dt \\ &= (\beta - \alpha^2/2)\ dt + \alpha\ dW_t. \end{aligned}$$

3.2. The vector version. We now introduce the vector version of Itô's formula. It is basically the same as before, but with worse notation.

Let $(W_t)_{t\geq 0}$ be a d-dimensional Wiener process, let $(\alpha_t)_{t\geq 0}$ be an adapted process valued in the space of $n \times d$ matrices, and let $(k_t)_{t\geq 0}$ be an adapted process valued in $\mathbb{R}^n$. We insist that

$$\int_0^t \sum_{i=1}^n \sum_{j=1}^d (\alpha_s^{(i,j)})^2 ds < \infty \text{ and } \int_0^t \sum_{i=1}^n |k_s^{(i)}| ds < \infty$$

almost surely for all $t \geq 0$. Now consider the n-dimensional Itô process $(X_t)_{t\geq 0}$ defined by

$$X_t = X_0 + \int_0^t \alpha_s dW_s + \int_0^t k_s ds,$$

interpreted component-wise as

$$X_t^{(i)} = X_0^{(i)} + \int_0^t \sum_{j=1}^d \alpha_s^{(i,j)} dW_s^{(j)} + \int_0^t k_s^{(i)} ds.$$

Now we are ready for the statement of the theorem:

THEOREM (Itô's formula, vector version). *Let $f : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}$ where $(t, x) \mapsto f(t, x)$ be continuously differentiable in the t variable and twice-continuously differentiable in the x variable. Then*

$$\boxed{df(t, X_t) = \frac{\partial f}{\partial t}(t, X_t)dt + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(t, X_t)\ dX_t^{(i)} + \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(t, X_t)\ d\langle X^{(i)}, X^{(j)}\rangle_t}$$

4. Girsanov's theorem

As we have seen in discrete time, the economic notion of an arbitrage-free market model is tied to the existence of a locally equivalent measure for which the discounted asset prices are martingales.

Recall that locally equivalent measures are related to positive martingales via the Radon–Nikodym theorem. Indeed, let $(\Omega, \mathcal{F}, \mathbb{P})$ be our probability space equipped with a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, and let $\mathbb{Q}$ be locally equivalent to $\mathbb{P}$ in the sense that the restrictions $\mathbb{P}|_t$ and $\mathbb{Q}|_t$ of $\mathbb{P}$ and $\mathbb{Q}$ to $\mathcal{F}_t$ are equivalent for each $t \geq 0$. Then, by the Radon–Nikodym theorem there exists a density

$$Z_t = \frac{d\mathbb{Q}|_t}{d\mathbb{P}|_t}$$

such that $(Z_t)_{t\in\mathbb{R}_+}$ is a strictly positive martingale on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Conversely, if $(Z_t)_{t\in\mathbb{R}_+}$ is a positive martingale we can define a locally equivalent measure $\mathbb{Q}$.

Motivated by above discussion, we aim to understand how martingales arise within the context of the Itô stochastic integration theory. Consider the stochastic process $(Z_t)_{t\in\mathbb{R}_+}$ given by

$$Z_t = e^{-\frac{1}{2}\int_0^t |\alpha_s|^2 ds + \int_0^t \alpha_s \cdot dW_s}$$

where $(W_t)_{t\in\mathbb{R}_+}$ is a d-dimensional Brownian motion and $(\alpha_t)_{t\in\mathbb{R}_+}$ is a d-dimensional adapted process with $\int_0^t |\alpha_s|^2 ds < \infty$.

This process is clearly positive. Furthermore, notice that by Itô's formula we have

$$dZ_t = Z_t \alpha_t \cdot dW_t$$

so that $(Z_t)_{t\in\mathbb{R}_+}$ is a local martingale, as it is a stochastic integral with respect to a Brownian motion.

What if $(Z_t)_{t\in\mathbb{R}_+}$ is a true martingale, and hence the density of a change of measure? What happens to the Brownian motion?

THEOREM (Cameron–Martin–Girsanov Theorem). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which a d-dimensional Brownian motion $(W_t)_{t\in\mathbb{R}_+}$ is defined, and let $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ be a filtration satisfying the usual conditions. Let*

$$Z_t = e^{-\frac{1}{2}\int_0^t |\alpha_s|^2 ds + \int_0^t \alpha_s \cdot dW_s}$$

and suppose $(Z_t)_{t\in\mathbb{R}_+}$ is a martingale. Define the locally equivalent measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ by the density process

$$\frac{d\mathbb{Q}|_t}{d\mathbb{P}|_t} = Z_t.$$

Then the d-dimensional process $(\hat{W}_t)_{t\in\mathbb{R}_+}$ defined by

$$\hat{W}_t = W_t - \int_0^t \alpha_s ds$$

is a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{Q})$.

Now, you may be asking yourself: When is the process $(Z_t)_{t\in\mathbb{R}_+}$ not just a local martingale, but a true martingale? An easy-to-check sufficient condition is given by:

THEOREM (Novikov's criterion). *If*

$$\mathbb{E}\left(e^{\frac{1}{2}\int_0^t |\alpha_s|^2 ds}\right) < \infty$$

for all $t \geq 0$ then the process $(Z_t)_{t\in\mathbb{R}_+}$ defined by $Z_t = e^{-\frac{1}{2}\int_0^t |\alpha_s|^2 ds + \int_0^t \alpha_s \cdot dW_s}$ is a true martingale.

5. Martingale representation theorems

In the introduction to this chapter, it was claimed that all continuous martingales are essentially stochastic integrals with respect to Brownian motion. In this section, we hopefully clear up the situation.

THEOREM (Martingale Representation Theorem). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which a d-dimensional Brownian motion $W = (W_t)_{t\in\mathbb{R}_+}$ is defined, and let the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ be the filtration generated by W.*

Let $L = (L_t)_{t\in\mathbb{R}_+}$ be a local martingale. Then there exists a unique adapted d-dimensional process $(\alpha_t)_{t\in\mathbb{R}_+}$ such that $\int_0^t |\alpha_s|^2 ds < \infty$ almost surely for all $t \geq 0$ and

$$L_t = L_0 + \int_0^t \alpha_s \cdot dW_s.$$

In particular, L is continuous.

Up to now, we have assumed that the filtration is generated by a Brownian motion. Even when it's not, we can say a lot about continuous local martingales:

THEOREM (Lévy's Characterization of Brownian Motion). *Let $(L_t)_{t\in\mathbb{R}_+}$ be a continuous d-dimensional local martingale such that*

$$\langle L^{(i)}, L^{(j)} \rangle_t = \begin{cases} t & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Then $(L_t)_{t\in\mathbb{R}_+}$ is a standard d-dimensional Brownian motion.

Proof. Fix a constant vector $\theta \in \mathbb{R}^d$ and let $i = \sqrt{-1}$. Consider

$$M_t = e^{i\ \theta \cdot L_t + |\theta|^2 t/2}.$$

By Itô's formula,

$$\begin{aligned} dM_t &= M_t \left(i\ \theta \cdot dL_t + \frac{|\theta|^2}{2} dt \right) - \frac{1}{2} M_t \sum_{m=1}^{d} \sum_{n=1}^{d} \theta^{(m)} \theta^{(n)} d\langle L^{(m)}, L^{(n)} \rangle_t \\ &= i\ M_t \theta \cdot dL_t \end{aligned}$$

and so $(M_t)_{t \in \mathbb{R}_+}$ is a continuous local martingale, as it is the stochastic integral with respect to a continuous local martingale. On the other hand, since $|M_t| = e^{|\theta|^2 t/2}$ and hence $\mathbb{E}(\sup_{s \in [0,t]} |M_s|) < \infty$ the process $(M_t)_{t \in \mathbb{R}_+}$ is a true martingale. Thus for all $0 \leq s \leq t$ we have

$$\mathbb{E}(M_t | \mathcal{F}_s) = M_s$$

which implies

$$\mathbb{E}(e^{i\ \theta \cdot (L_t - L_s)} | \mathcal{F}_s) = e^{-|\theta|^2 (t-s)/2}.$$

The above equation implies that the increment $L_t - L_s$ has the $N_d(0, (t-s)I)$ distribution and is independent of $\mathcal{F}_s$. □

Example. Let $(L_t)_{t \in \mathbb{R}_+}$ be a continuous real-valued local martingale, and suppose there is an adapted process $(\alpha_t)_{t \in \mathbb{R}_+}$ such that $\langle L \rangle_t = \int_0^t \alpha_s^2\ ds$ where $\alpha_t > 0$ almost surely. $W_t = \int_0^t \frac{dL_s}{\alpha_s}$. Note that $(W_t)_{t \in \mathbb{R}_+}$ is a Brownian motion by Lévy's characterization since

$$\langle W \rangle_t = \int_0^t \frac{d\langle L \rangle_s}{\alpha_s^2} = t.$$

Hence, $(W_t)_{t \in \mathbb{R}_+}$ is a Brownian motion such that

$$L_t = L_0 + \int_0^t \alpha_s\ dW_s$$

for all $t \geq 0$.

As another application of Lévy's characterization of Brownian motion, we can attempt a proof of Girsanov's theorem:

Proof of Girsanov's theorem. Let

$$Z_t = e^{-\frac{1}{2} \int_0^t |\alpha_s|^2 ds + \int_0^t \alpha_s \cdot dW_s}$$

for a d-dimensional adapted process $(\alpha_t)_{t \in \mathbb{R}_+}$ and a d-dimensional Brownian motion $(W_t)_{t \in \mathbb{R}_+}$, and suppose $(Z_t)_{t \in \mathbb{R}_+}$ is a true martingale. Let

$$\hat{W}_t = W_t - \int_0^t \alpha_s\ ds.$$

Note that $(Z_t\hat{W}_t)_{t\in\mathbb{R}_+}$ is a local martingale for $\mathbb{P}$, since by Itô's formula:

$$\begin{aligned} d(Z_t\hat{W}_t^{(i)}) &= \hat{W}_t^{(i)}dZ_t + Z_t d\hat{W}_t^{(i)} + d\langle Z, \hat{W}^{(i)}\rangle_t \\ &= Z_t\hat{W}_t^{(i)}\alpha_t \cdot dW_t + Z_t dW_t^{(i)} \end{aligned}$$

It now follows that $(\hat{W}_t)_{t\in\mathbb{R}_+}$ is a local martingale for $\mathbb{Q}$, where $\frac{d\mathbb{Q}|_t}{d\mathbb{P}|_t} = Z_t$. But since

$$\langle \hat{W}^{(i)}, \hat{W}^{(j)}\rangle_t = \begin{cases} t & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

then $(\hat{W}_t)_{t\in\mathbb{R}_+}$ is a Brownian motion of for $\mathbb{Q}$. □

CHAPTER 4

Black–Scholes model and generalizations

We now return to the main theme of these lecture, models of financial markets. We now have the tools to discuss the continuous time case, at least when the asset prices are continuous processes.

1. The set-up

As before, our market model consists of a $d+1$-dimensional stochastic processes $(B, S) = (B_t, S_t)_{t\in\mathbb{R}_+}$ representing the asset prices. This process will be defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ satisfying the usual conditions.

We will make the following now-familiar assumptions.

ASSUMPTION. The stochastic process (B, S) is assumed to be is an Itô process adapted to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$.

ASSUMPTION. Asset 0 is a numéraire so that $B_t > 0$ almost surely for all $t \in \mathbb{R}_+$.

As before, the investor's controls consist of the $d+1$-dimensional process $(\phi_t, \pi_t)_{t\in\mathbb{R}_+}$, where ϕ_t and corresponds to the number of shares of asset 0 held at time t, while $\pi_t^{(i)}$ corresponds to the number of shares of asset $i \in \{1, \dots, d\}$ held at time t. We will use the notation $\pi_t = (\pi_t^1, \dots, \pi_t^d)$.

The wealth X_t at time t then satisfies the pair of equations

$$\boxed{\begin{array}{rclr} X_t &=& \phi_t B_t + \pi_t \cdot S_t & \text{the budget constraint} \\ dX_t &=& \phi_t dB_t + \pi_t \cdot dS_t & \text{the self-financing condition} \end{array}}$$

As usual, we can use the budget contraint to solve for ϕ_t, and plug this expression into the self-financing condition to get

$$dX_t = (X_t - \pi_t \cdot S_t)\frac{dB_t}{B_t} + \pi_t \cdot dS_t$$

By Itô's formula, we have

$$\begin{aligned} d\left(\frac{X_t}{B_t}\right) &= X_t\left(-\frac{dB_t}{B_t^2} + \frac{d\langle B\rangle_t}{B_t^3}\right) + \frac{dX_t}{B_t} - \frac{d\langle X, B\rangle_t}{B_t^2} \\ &= \sum_{i=1}^d \pi_t^{(i)}\left[S_t^{(i)}\left(-\frac{dB_t}{B_t^2} + \frac{d\langle B\rangle_t}{B_t^3}\right) + \frac{dS_t}{B_t} - \frac{d\langle S^{(i)}, B\rangle_t}{B_t^2}\right] \\ &= \pi_t \cdot d\left(\frac{S_t}{B_t}\right). \end{aligned}$$

As usual, we express everything in terms of asset 0 prices

$$\tilde{S}_t = \frac{S_t}{B_t} \text{ and } \tilde{X}_t = \frac{X_t}{B_t}$$

so that

$$\boxed{\tilde{X}_t = \tilde{X}_0 + \int_0^t \pi_s \cdot d\tilde{S}_t}$$

2. Admissible strategies

In order to make sense of the stochastic integral defining the discounted wealth, we need to impose the integrablity condition

$$\int_0^t \sum_{i=1}^d (\pi_s^i)^2 d\langle \tilde{S}^i \rangle_s < \infty$$

almost surely for all $t \geq 0$.

However, in moving from discrete to continuous time, we have to be careful. We will now see that this condition isn't strong enough to make our economic analysis interesting.

Example. (A doubling strategy.) This example is intended to provide motivation for restricting the class of trading strategies that we will consider in these lectures.

The problem with continuous time is that events that will happen eventually can be made to happen at any finite time by speeding up the clock. In particular, we will now construct a real-valued adapted process $(\alpha_t)_{t\in[0,T]}$ such that $\int_0^T \alpha_s^2 ds < \infty$ almost surely, but

$$\int_0^T \alpha_s dW_s = K$$

almost surely, where $(W_t)_{t\in[0,T]}$ is a standard scalar Brownian motion, and $T > 0$ and $K > 0$ are real constants.

Let $f : [0, T) \to \mathbb{R}_+$ be a strictly increasing, continuous function such that $f(0) = 0$ and $\lim_{t\to T} f(t) = \infty$. Note in particular that we assume that $f'(t) > 0$ for $t \in [0, T)$ and there exists an inverse function $f^{-1} : \mathbb{R} \to [0, T)$ such that $f \circ f^{-1}(u) = u$. For instance, to be explicit, we may take $f(t) = \frac{t}{T-t}$ and $f^{-1}(u) = \frac{Tu}{1+u}$.

Now define a process $(Z_u)_{u\in\mathbb{R}_+}$ by

$$Z_u = \int_0^{f^{-1}(u)} (f'(s))^{1/2} dW_s$$

Note that the quadratic variation is

$$\begin{aligned}\langle Z \rangle_u &= \int_0^{f^{-1}(u)} f'(s) ds \\ &= f(f^{-1}(u)) - f(0) \\ &= u\end{aligned}$$

so by Lévy's characterization $(Z_u)_{u\in\mathbb{R}_+}$ is a Brownian motion. Define the stopping time τ by

$$\tau = \inf\{u \geq 0, Z_u = K\}.$$

Since $(Z_u)_{u\in\mathbb{R}_+}$ is a Brownian motion, we have $\tau < \infty$ almost surely.[1] Now let

$$\alpha_s = (f'(s))^{1/2}\mathbb{1}_{\{s\leq f^{-1}(\tau)\}}$$

and

$$M_t = \int_0^t \alpha_s dW_s$$

Note that since

$$\int_0^T \alpha_s^2 ds = \int_0^{f^{-1}(\tau)} f'(s)ds = \tau < \infty$$

the stochastic integral is well-defined. The strange fact is that $(M_t)_{t\in[0,T]}$ is a local martingale with $M_0 = 0$, but $M_T = Z_\tau = K$ almost surely.

We see that integrand $(\alpha_s)_{s\in[0,T]}$ corresponds to an gambler starting at noon with £0, employing a doubling strategy (with borrowed money) at a quicker and quicker pace, until finally he gains £K almost surely before the clock strikes one o'clock. This situation is rather unrealistic, particularly since the gambler must go arbitrarily far into debt in order to secure the £K winning. Indeed, if such strategies were a good model for investor behaviour, we all could be much richer by just spending some time trading over the internet.

The above discussion shows that the natural integrability condition

$$\int_0^t \sum_{i=1}^d (\pi_s^i)^2 d\langle \tilde{S}^i\rangle_s < \infty \text{ almost surely}$$

is not really sufficient for our needs.

We might choose to impose the stronger condition

$$\mathbb{E}\left(\int_0^t \sum_{i=1}^d (\pi_s^i)^2 d\langle \tilde{S}^{(i)}\rangle_s\right) < \infty$$

for all $t \geq 0$. In this case, we would be able to use the L^2 theory, and in particular, the stochastic integrals against Brownian motions would be martingales. However, we usually do not impose this extra integrability condition because in finance we are often computing expectations under an equivalent measure. That is, even if $\mathbb{P}$ and $\mathbb{Q}$ are equivalent and Y is a positive random variable such that $\mathbb{E}^{\mathbb{P}}(Y) < \infty$, it doesn't necessarily follow that $\mathbb{E}^{\mathbb{Q}}(Y) < \infty$. (Can you find an example?) So instead, we usually insist that the investor cannot go arbitrarily far into debt.

Definition. Let (B, S) be a market model. A strategy $(\pi_t)_{t\in\mathbb{R}_+}$ is *admissible* if and only if there is a constant $C > 0$ such that

$$\int_0^t \pi_s \cdot d\tilde{S}_s > -C \text{ almost surely}$$

[1] To see why, let $Z_u^{(n)} = \sqrt{n}Z_{u/n}$. One can check that that the proces $(Z_u^{(n)})_{u\in\mathbb{R}_+}$ is also a standard Brownian motion. Hence

$$\mathbb{P}(\sup_{u\geq 0} Z_u > K) = \mathbb{P}(\sup_{u\geq 0}\sqrt{n}Z_{u/n} > K) = \mathbb{P}(\sup_{u\geq 0} Z_u > K/\sqrt{n}) \to \mathbb{P}(\sup_{u\geq 0} Z_u > 0).$$

But it is easy to see that $\mathbb{P}(\sup_{u\geq 0} Z_u > 0) = 1$.

for all $t \geq 0$, where $\tilde{S}_t = \frac{S_t}{B_t}$.

Note that the doubling strategy is not admissible, since the investor now has only a finite credit line. However, a *suicide strategy*, that is, a doubling strategy in which the object is to *lose* a fixed amount K by time T, is admissible.

3. Arbitrage and equivalent martingale measures

To see that our restriction to admissible strategies is reasonable, let's now consider continuous-time arbitrage theory.

DEFINITION. An admissible strategy $(\pi_t)_{t\in\mathbb{R}_+}$ is an *arbitrage* if there is a (non-random) time $T > 0$ such that

$$\begin{aligned} \mathbb{P}\left(\int_0^T \pi_s \cdot d\tilde{S}_s \geq 0\right) &= 1 \\ \mathbb{P}\left(\int_0^T \pi_s \cdot d\tilde{S}_s > 0\right) &> 0. \end{aligned}$$

DEFINITION. A probability measure $\mathbb{Q}$, locally equivalent to $\mathbb{P}$, such that $\tilde{S} = (\tilde{S})_{t\in\mathbb{R}_+}$ is a local martingale is called a *equivalent martingale measure*.

The following theorem will serve the role of the first fundamental theorem of asset pricing in continuous time.

THEOREM. *If there exists an equivalent martingale measure, then the market model has no arbitrage.*

REMARK. Note that the theorem doesn't say that no-arbitrage implies the existence of an equivalent martingale measure. Indeed, our notion of arbitrage is too strong. The 'correct' notion of arbitrage is called 'free-lunch-with-vanishing-risk,' but it is outside the scope of these lectures. See the recent book of Delbaen and Schachermayer *The Mathematics of Arbitrage* for an account of the modern theory.

PROOF. Suppose $(\pi_t)_{t\in\mathbb{R}_+}$ is admissible and let

$$\tilde{X}_t = \int_0^t \pi_s \cdot d\tilde{S}_s.$$

By the definition, there exists a constant $C > 0$ such that $\tilde{X}_t > -C$ $\mathbb{P}-$ almost surely. Since $\mathbb{P}$ and $\mathbb{Q}$ are locally equivalent, then $\tilde{X}_t > -C$ $\mathbb{Q}-$ almost surely. But under $\mathbb{Q}$, the process $(\tilde{S}_t)_{t\in\mathbb{R}_+}$ is a local martingale. But a local martingale that is bounded from below is necessarily a supemartingale. (See Problem 3.5) In particular, we have the following inequality

$$\mathbb{E}^{\mathbb{Q}}(\tilde{X}_t) \leq \tilde{X}_0 = 0.$$

for all $t \geq 0$.

Now suppose that there is a $T > 0$ such that $\tilde{X}_T \geq 0$ $\mathbb{P}$-almost surely. Then $\tilde{X}_T \geq 0$ $\mathbb{Q}$-almost surely. But since $\mathbb{E}^{\mathbb{Q}}(\tilde{X}_T) \leq 0$, we conclude that $\tilde{X}_T = 0$ $\mathbb{Q}$-almost surely, which implies $\tilde{X}_T = 0$ $\mathbb{P}$-almost surely. □

4. Contingent claims and market completeness

As before, given a market model (B, S), we can introduce a contingent claim. Recall that a European contingent claim maturing at a time $T > 0$ is modelled as random variable ξ that is $\mathcal{F}_T$-measurable. Let $\mathcal{Q}$ be the set of equivalent martingale measures, and which we shall assume is not empty. Hence, there is no arbitrage in the market.

THEOREM. *Let $(\xi_t)_{t\in[0,T]}$ be a process such that $\xi_T = \xi$. The augmented market $(B_t, S_t, \xi_t)_{t\in[0,T]}$ is free of arbitrage if there exists an equivalent martingale measure $\mathbb{Q} \in \mathcal{Q}$ such that*

$$\xi_t = \mathbb{E}^{\mathbb{Q}}\left(\frac{B_t}{B_T}\xi \middle| \mathcal{F}_t\right)$$

for all $t \in [0, T]$.

PROOF. There is no arbitrage if there exists an equivalent measure $\mathbb{Q}$ such that $(\tilde{S}_t, \tilde{\xi}_t)_{t\in[0,T]}$ is a local martingale for $\mathbb{Q}$, where $\tilde{\xi}_t = \frac{\xi_t}{B_t}$. But $(\tilde{\xi}_t)_{t\in[0,T]}$ is a martingale by construction. □

And just as before, we are interested in replicating contingent claims by trading in the market.

DEFINITION. A (European) contingent claim with payout ξ_T maturing at time $T > 0$ is *attainable* if and only if there exists a constant $x \in \mathbb{R}$ and an admissible strategy $(\pi)_{t\in[0,T]}$ such that

$$\tilde{\xi}_T = x + \int_0^T \pi_s \cdot d\tilde{S}_s$$

where $\tilde{\xi}_T = \frac{\xi_T}{B_T}$.

A market is *complete* if for every bounded $\mathcal{F}_T$-measurable random variable $\tilde{\xi}_T$, there exists x, π such that

$$\tilde{X}_t = x + \int_0^t \pi_s \cdot d\tilde{S}_s$$

defines a bounded discounted wealth process $\tilde{X}$ such that $X_T = \xi_T$.

The following is a version of the second fundamental theorem of asset pricing for this continuous time setting.

THEOREM. *Suppose that there exists an equivalent martingale measure $\mathbb{Q}$ and the market is complete. Then $\mathbb{Q}$ is the unique equivalent martingale measure.*

PROOF. Let $\mathbb{Q}'$ be another equivalent martingale measures and let $\tilde{\xi}_T$ be an arbitrary bounded $\mathcal{F}_T$-measurable random variable. By assumption ξ_T is attainable by a wealth process X, and since $\tilde{X}$ is a bounded, it is a martingale for both $\mathbb{Q}$ and $\mathbb{Q}'$.

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}'}(\tilde{\xi}_T) &= \mathbb{E}^{\mathbb{Q}'}(\tilde{X}_T) \\ &= x = \mathbb{E}^{\mathbb{Q}}(\tilde{\xi}_T)\end{aligned}$$

Hence, for all bounded random variables $\tilde{\xi}_T$, the inequality $\mathbb{E}^{\mathbb{P}}[(Z_T - Z'_T)\tilde{\xi}_T] \geq 0$ holds, where $Z_T = \frac{d\mathbb{Q}_T}{d\mathbb{P}_T}$ and $Z'_T = \frac{d\mathbb{Q}'_T}{d\mathbb{P}_T}$. Letting $\tilde{\xi} = \mathbb{1}_{\{Z_T < Z'_T\}}$ shows that $\mathbb{P}(Z_T \geq Z'_T) = 1$. Symmetry completes the argument shows $Z_T = Z'_T$ a.s. □

At this stage we're a bit too general to say anything interesting. Hence, it is wise to introduce more notation to flesh out the story...

5. The set-up revisited

In this section, we revisit the set-up of the continuous time market model (B, S). We will make one extra assumption, which is standard, but not really necessary:

ASSUMPTION. Asset 0 is risk-free in the sense that

$$\langle B \rangle_t = 0 \text{ almost surely.}$$

The point of this section is to set notation that will be used for the remaining lectures, and to interpret the existence of an equivalent martingale measure via Girsanov's theorem.

Asset 0, which we now assume is risk-free, will be intepreted as a bank account. Its dynamics are given by

$$dB_t = r_t B_t dt$$

for an adapted process $(r_t)_{t \geq 0}$ which can be interpreted as the spot interest rate. The above random ordinary differential equation has the solution

$$B_t = B_0 e^{\int_0^t r_s ds}.$$

Assets $1, \ldots, d$ are assumed to have dynamics given by

$$dS_t^{(i)} = S_t^{(i)} \left(\mu_t^{(i)} \; dt + \sum_{j=1}^n \sigma_t^{(i,j)} dW_t^{(j)} \right)$$

for some adapted process $(\mu_t^{(i)})_{t \in \mathbb{R}_+}$ and $(\sigma_t^{(i,j)})_{t \in \mathbb{R}_+}$, and independent Brownian motions $(W_t^{(j)})_{t \in \mathbb{R}_+}$. The random variable $\mu_t^{(i)}$ can be considered the instantaneous drift of the return on stock i, while the random variable

$$\left(\sum_{j=1}^n (\sigma_t^{(i,j)})^2 \right)^{1/2}$$

is its instantaneous volatility. In vector notation, these equations can be written as

$$dS_t = \text{diag}(S_t)(\mu_t dt + \sigma_t dW_t)$$

where $S_t = (S_t^{(1)}, \ldots, S_t^{(d)})$, and the $\mathbb{R}^d$-valued random variables μ_t and W_t, and $d \times n$ matrix-valued random σ_t defined similarly. Here we are using the notation

$$\text{diag}(s_1, \ldots, s_d) = \begin{pmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & s_d \end{pmatrix}$$

The dynamics of the discounted stock price are given by

$$d\tilde{S}_t = \text{diag}(\tilde{S}_t)((\mu_t - r_t \mathbf{1})dt + \sigma_t dW_t)$$

where $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^d$.

Now we come to the two theorems of this section: The first is a reasonably easy-to-check sufficient condition to ensure no-arbitrage.

THEOREM. *Suppose there exists a predictable process* λ *such that*

$$\sigma_t \lambda_t = \mu_t - r_t \mathbf{1}$$

almost surely for all $t \geq 0$. *If the process* λ *is such that*

$$Z_t = e^{-\frac{1}{2}\int_0^t |\lambda_s|^2 \, ds - \int_0^t \lambda_s \cdot dW_s}.$$

defines a martingale (for instance if Novikov's criterion

$$\mathbb{E}^{\mathbb{P}}(e^{\frac{1}{2}\int_0^t |\lambda_s|^2 \, ds}) < \infty$$

holds for all $t \geq 0$*), then the market has no arbitrage.*

PROOF. By Girsanov's theorem, the locally equivalent measure $\mathbb{Q}$ whose density process is Z is such that the process $\hat{W}$ given by

$$\hat{W}_t = W_t + \int_0^t \lambda_s \, ds$$

is a Brownian motion for $\mathbb{Q}$.

Notice then that the dynamics of the discounted stock price are given by

$$d\tilde{S}_t = \operatorname{diag}(\tilde{S}_t)\sigma_t d\hat{W}_t,$$

so that $\tilde{S}$ is a local martingale for $\mathbb{Q}$. Hence $\mathbb{Q}$ is an equivalent martingale measure and there is no arbitrage in this market. □

REMARK. The n-dimensional random vector λ_t is a generalization of the Sharpe ratio. The process $\lambda = (\lambda_t)_{t \in \mathbb{R}_+}$ is often called the *market price of risk*, since it measures in some sense the excess return of the stocks per unit of volatility.

The following theorem is a sufficient condition for completeness:

THEOREM. *Take as given the hypotheses of the previous theorem, and further suppose* $n = d$ *and* $\sigma_t(\omega)$ *is invertible for all* (t, ω)*, so that* $\lambda_t = \sigma_t^{-1}(\mu_t - r_t \mathbf{1})$*. Let* $d\hat{W}_t = dW_t + \lambda_t \, dt$*. If the process* $\hat{W}$ *generates the filtration, then the market is complete.*

PROOF. Let $\mathbb{Q}$ be the equivalent martingale measure in the proof of the previous theorem. Fix a bounded $\mathcal{F}_T$-measurable random variable $\tilde{\xi}_T$, and let

$$\tilde{\xi}_t = \mathbb{E}^{\mathbb{Q}}(\tilde{\xi}_T | \mathcal{F}_t).$$

so that $(\tilde{\xi}_t)_{t \in [0,T]}$ is a bounded $\mathbb{Q}$-martingale. Since $\hat{W}$ is a Brownian motion under $\mathbb{Q}$, and generates the filtration, the martingale representation theorem asserts the existence an adapted process $(\alpha_t)_{t \in [0,T]}$ such that

$$\tilde{\xi}_t = \mathbb{E}^{\mathbb{Q}}(\tilde{\xi}_T) + \int_0^t \alpha_s \cdot d\hat{W}_s.$$

On the other hand, we can write the discounted stock dynamics as

$$d\tilde{S}_t = \operatorname{diag}(\tilde{S}_t)\sigma_t d\hat{W}_t.$$

By taking $x = \mathbb{E}^{\mathbb{Q}}(\tilde{\xi}_T)$ and $\pi_t = \operatorname{diag}(S_t)^{-1}(\sigma_t^{\mathrm{T}})^{-1}\alpha_t$, we see that

$$\tilde{\xi}_t = x + \int_0^t \pi_s \cdot d\tilde{S}_s,$$

and hence the claim with payout $\xi_T = X_T$ is attainable, and the discounted wealth process $\tilde{X} = \tilde{\xi}$ is bounded. Hence, the market is complete. □

If we consider the equation $\sigma_t \lambda_t = \mu_t - r_t \mathbf{1}$ where σ_t is an $d \times n$ matrix, one expects from the rules of linear algebra for there to be no solution if $n < d$, exactly one solution if $n = d$, and many solutions if $n > d$. Of course, this is not theorem, just a rule of thumb. Financially, the rule of thumb becomes:

$n < d$ '$\Rightarrow$'	The market has arbitrage.
$n = d$ '$\Rightarrow$'	The market has no arbitrage and is complete.
$n > d$ '$\Rightarrow$'	The market has no arbitrage and is incomplete.

We now have a sufficient condition that the market model is complete. However, at this stage we can only assert the existence of a replicating strategy for a given claim, but we do not yet know how to actually compute it. This problem is the subject of the next section.

6. Markovian markets

Now that we have our two main structural theorems in the context of a market with continuous asset prices, we are still left with the question: How do you price and hedge contingent claims?

The first step is to pose a model for the asset prices $(B_t, S_t)_{t \in \mathbb{R}_+}$. A good model should give a reasonable statistical fit to the actual market data. See figure 1 below.

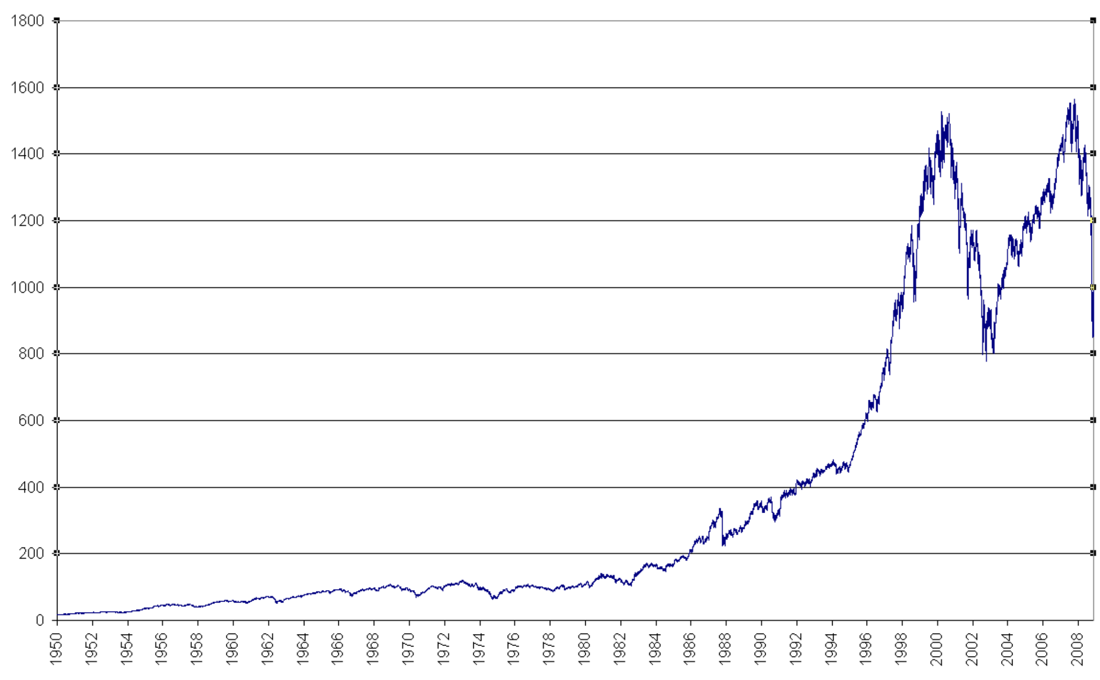


FIGURE 1. Graph of the Standard & Poor's 500 stock index 1950-2008. Data taken from `http://finance.yahoo.com`

Furthermore, a *useful* model is one in which the prices and hedges of contingent claims can be computed reasonably easily. In this section, we will study models in which the asset prices are Markov processes. These models are useful in the above sense, though there seems to be some controversy over how well they fit actual market data.

Now suppose that the $d+1$ assets have Itô dynamics which can be expressed as

$$\begin{aligned} dB_t &= B_t\, r(t, S_t)\, dt \\ dS_t &= \operatorname{diag}(S_t)(\mu(t, S_t)dt + \sigma(t, S_t)dW_t) \end{aligned}$$

where the nonrandom functions $r : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$, $\mu : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^{d\times n}$ are given. Notice that this is a special case of the set-up of the last section, as now

$$r_t(\omega) = r(t, S_t(\omega)), \quad \mu_t(\omega) = \mu(t, S_t(\omega)), \text{ and } \sigma_t(\omega) = \sigma(t, S_t(\omega)).$$

In this special situation, the asset prices $(S_t)_{t\in\mathbb{R}_+}$ are a d-dimensional Markov process.

The next theorem says how to find the no-arbitrage price *and* the replicating strategy for a contingent claim maturing at time T with payout

$$\xi = g(S_T)$$

for some non-random function $g : \mathbb{R}^d \to \mathbb{R}$.

THEOREM. *Assume there exists an equivalent martingale measure.*
Suppose the function $V : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ satifies the partial differential equation

$$\begin{aligned} &\frac{\partial V}{\partial t} + \sum_{i=1}^{d} r S^i \frac{\partial V}{\partial S^i} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} a_{i,j} S^i S^j \frac{\partial^2 V}{\partial S^i \partial S^j} = rV \\ &V(T, S) = g(S) \end{aligned}$$

where $a = \sigma\sigma^{\mathrm{T}}$, and where all functions in the PDE are evaluated at the same point $(t, S) \in [0, T) \times \mathbb{R}^d$.
If

$$\xi_t = V(t, S_t)$$

then there is no arbitrage in the augmented market $(B_t, S_t, \xi_t)_{t\in[0,T]}$.
Furthermore, if V is non-negative, then the claim with payout $\xi = g(S_T)$ is attainable with initial capital

$$\xi_0 = V(0, S_0)$$

and the replicating strategy

$$\pi_t = \text{grad } V(t, S_t) = \left(\frac{\partial V}{\partial S^1}(t, S_t), \ldots, \frac{\partial V}{\partial S^d}(t, S_t)\right).$$

REMARK. The above theorem says that if the market model is Markovian, the price of a claim contingent on the future risky asset prices can be written as a deterministic function V of the current market prices. Furthermore, the pricing function V can be found by solving a certain linear parabolic partial differential equation[2] with terminal data to match the payout of the claim. Solving this equation may be difficult to do by hand, but it can usually be done by computer if the dimension d is reasonably small. And most importantly for the banker selling such a contingent claim: the replicating portfolio π_t can be calculated as the

[2]sometimes called the Feynman–Kac PDE. If $r = 0$, the PDE reduces to the (backward) Kolmogorov equation.

gradient of the pricing function V with respect to the spatial variables, evaluated at time t and current price S_t.

PROOF. Let $\mathbb{Q}$ be an equivalent martingale measure, i.e. a measure under which $\tilde{S}$ is a local martingale. By Itô's formula, we have

$$
\begin{aligned}
d\tilde{\xi}_t &= d\left(\frac{V(t,S_t)}{B_t}\right) \\
&= \frac{1}{B_t}\left(\frac{\partial V}{\partial t} + \sum_{i=1}^{d} \frac{\partial V}{\partial S^i} dS_t^i + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2 V}{\partial S^i \partial S^j} d\langle S^i, S^j\rangle_t\right) - V(t,S_t)\frac{dB_t}{B_t^2} \\
&= \frac{1}{B_t}\left(\frac{\partial V}{\partial t} + \sum_{i=1}^{d} \mu^i S_t^i \frac{\partial V}{\partial S^i} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{n} \sigma^{ik}\sigma^{jk} S_t^i S_t^j \frac{\partial^2 V}{\partial S^i \partial S^j} - rV\right) dt \\
&\quad + \frac{1}{B_t}\sum_{i=1}^{d}\sum_{j=1}^{n} \frac{\partial V}{\partial S^i} S^i \sigma^{ij} dW_t^j \\
&= \sum_{i=1}^{d} \frac{\partial V}{\partial S^i}\frac{S^i}{B_t}\left[(\mu^i - r)dt + \sum_{j=1}^{n}\sigma^{ij} dW_t^j\right] \\
&= \text{grad } V \cdot d\tilde{S}_t.
\end{aligned}
$$

Hence $(\tilde{\xi}_t)_{t\in[0,T]}$ is a local martingale for $\mathbb{Q}$.

Furthermore, if V is non-negative, ξ can be attained by the announced admissible trading strategy $\pi_t = \text{grad } V(t, S_t)$. □

7. The Black–Scholes model, PDE, and formula

In this section, we will consider the simplest possible Markovian model of the type studied in section 6. Consider the case of a market with two assets. We will assume that all coefficients are constant, so the price dynamics are given by the pair of equations

$$
\begin{aligned}
dB_t &= B_t \, r \, dt \\
dS_t &= S_t(\mu \, dt + \sigma dW_t)
\end{aligned}
$$

for real constants r, μ, σ where $\sigma > 0$. This is often called the *Black–Scholes* model.

We've seen before that since $\lambda = (\mu - r)/\sigma$ is constant, there exists a locally equivalent measure $\mathbb{Q}$ such that the process defined by $\hat{W}_t = W_t + \lambda t$ is a Brownian motion. In particular, the market has no arbitrage. Furthermore, if the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is the natural filtration of the asset prices $(B_t, S_t)_{t\in\mathbb{R}_+}$, the model is complete.

We are interested in pricing and hedging a European contingent claim with payout $\xi_T = g(S_T)$. As we've seen, there are two ways of doing this.

7.1. Pricing by expectations.

. From Section 4, there is no arbitrage if

$$
\xi_t = \mathbb{E}^{\mathbb{Q}}\left[\frac{B_t}{B_T} g(S_T) | \mathcal{F}_t\right]
$$

assuming the expectation exists. In this simple case, the prices of traded assets can be written explicitly:

$$B_t = B_0 e^{rt} \text{ and } S_t = S_0 e^{(\mu-\sigma^2/2)t+\sigma W_t} = S_0 e^{(r-\sigma^2/2)t+\sigma \hat{W}_t}$$

and hence

$$\begin{aligned}
\xi_t &= \mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t)} g\left(S_0 e^{(r-\sigma^2/2)T+\sigma\hat{W}_T}\right)|\mathcal{F}_t\right] \\
&= \mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t)} g\left(S_t e^{(r-\sigma^2/2)(T-t)+\sigma(\hat{W}_T-\hat{W}_t)}\right)|\mathcal{F}_t\right] \\
&= e^{-r(T-t)} \int_{-\infty}^{\infty} g\left(S_t e^{(r-\sigma^2/2)(T-t)+\sigma\sqrt{T-t}z}\right) \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz.
\end{aligned}$$

In section 5, we argued that, assuming the filtration is generated by S, the martingale representation theorem asserts the existence of a process π such that

$$e^{-rT} g(S_T) = \mathbb{E}^{\mathbb{Q}}[e^{-rT} g(S_T)] + \int_0^T \pi_s d\tilde{S}_s$$

Unfortunately, we do not know how to compute π...

7.2. Pricing and hedging by PDE.. From the general results of section 6, we can solve the *Black–Scholes* PDE

$$\boxed{\begin{array}{l} \dfrac{\partial V}{\partial t} + rS\dfrac{\partial V}{\partial S} + \dfrac{1}{2}\sigma^2 S^2 \dfrac{\partial^2 V}{\partial S^2} = rV \\ V(T,S) = g(S) \end{array}}$$

If $\xi_t = V(t, S_t)$ then $\xi_T = g(S_T)$ and $(\tilde{\xi}_t)_{t\in[0,T]}$ is a local martingale, hence ξ_t is a no-arbitrage price for the claim. Furthermore, assuming V is bounded from below, we see that $\pi_t = \frac{\partial V}{\partial S}(t, S_t)$ is an admissible replicating portfolio

$$e^{-rT} g(S_T) = V(0, S_0) + \int_0^T \frac{\partial V}{\partial S}(s, S_s) d\tilde{S}_s.$$

Because of its central importance in the theory, the quantity $\frac{\partial V}{\partial S}$ is given a special name, the *delta*, of the claim.[3]

Remark. In nearly all cases of interest, the two approaches yield the same answer. Here is a sufficient condition: Suppose there exists positive contants C and p such that

$$|V(t,S)| \le C(1 + S^p)$$

for all $(t, S) \in [0, T] \times \mathbb{R}_+$, then

$$V(t, S_t) = \mathbb{E}^{\mathbb{Q}}[e^{-r(T-t)} g(S_T)|\mathcal{F}_t].$$

It is worth pointing out, however, that the above sufficient condition is specific to the Black–Scholes model.

[3]The name delta is inspired by the notation of the original Black–Scholes paper. In finance, there is a whole list if quantities, the delta, theta, gamma, vega, etc. which describe the sensitivities of a pricing formula with respect to some of its parameters. These quantities are collectively knowns as the *greeks*.

We can easily apply this result to a specific payout function g. For instance, introduce a European call option maturing at T with strike K. The payout is $(S_T - K)^+$. There is no arbitrage if the time t price $C_t(T,K)$ of the call is given by

$$C_t(T,K) = \mathbb{E}^{\mathbb{Q}}[e^{-r(T-t)}(S_T - K)^+ | \mathcal{F}_t]$$

yielding the Nobel-prize-winning *Black–Scholes formula*:

$$\boxed{\begin{aligned} C_t(T,K) &= S_t\ \Phi\left(\frac{\log(S_t/K)}{\sigma\sqrt{T-t}} + (r/\sigma + \sigma/2)\sqrt{T-t}\right) \\ &\quad -e^{-r(T-t)}K\ \Phi\left(\frac{\log(S_t/K)}{\sigma\sqrt{T-t}} + (r/\sigma - \sigma/2)\sqrt{T-t}\right). \end{aligned}}$$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$ is the standard normal distribution function. Notice that in this case, the delta, i.e. the hedging portfolio, is given by

$$\pi_t = \Phi\left(\frac{\log(S_t/K)}{\sigma\sqrt{T-t}} + (r/\sigma + \sigma/2)\sqrt{T-t}\right).$$

7.3. Volatility estimation. What made this formula so popular after its publication in 1973 is the fact that the right-hand-side depends only on six quantities: the current calendar time t, the option's maturity time T, the option's strike K, the spot interest rate r, the underlying stock's price S_t at time t, and a volatility parameter σ. Of these six numbers, only the volatility parameter is neither specified by the option contract nor quoted in the market.

To use the Black–Scholes formula to find the price of real call options, one must first estimate the volatility σ. One possibility is to is to collect $n+1$ historical price observations at times $t_0 < \ldots < t_n$. Then the random variables $Y_i = \log\left(\frac{S_{t_i}}{S_{t_{i-1}}}\right)$ then has distribution

$$\begin{aligned} Y_i &= (\mu - \sigma^2/2)(t_i - t_{i-1}) + \sigma(W_{t_i} - W_{t_{i-1}}) \\ &\sim N(\nu\Delta t_i, \sigma^2\Delta t_i) \end{aligned}$$

where $\nu = \mu - \sigma^2/2$ and $\Delta t_i = t_i - t_{i-1}$. The maximum likelihood estimators are then

$$\hat{\nu} = \frac{1}{t_n - t_0}\sum_{i=1}^{n} Y_i$$

and

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} \frac{(Y_i - \hat{\nu}\Delta t_i)^2}{\Delta t_i}.$$

If one was to truly believe that the stock price was a *geometric Brownian motion*, that is, of the form $S_t = S_0 e^{\nu t + \sigma W_t}$, then one could insert the value $\hat{\sigma^2}$ into the Black–Scholes formula to obtain the price of a call option. Notice that we have done the statistics under the *objective* measure $\mathbb{P}$, not the equivalent martingale measure $\mathbb{Q}$.

7.4. Implied volatility. A completely different approach to find the volatility parameter is to observe the price $C_t(T,K)$ from the market, and then try to work out which σ to put into the Black–Scholes to get the right price.

Let the nonrandom function $\mathrm{BS} : \mathbb{R}_+ \times \mathbb{R}_+ \to [0,1)$ be defined by

$$\mathrm{BS}(v,m) = \begin{cases} \Phi\big(-\frac{\log m}{\sqrt{v}} + \frac{\sqrt{v}}{2}\big) - m\Phi\big(-\frac{\log m}{\sqrt{v}} - \frac{\sqrt{v}}{2}\big) & \text{if } v > 0 \\ (1-m)^+ & \text{if } v = 0. \end{cases}$$

Then Black–Scholes formula says that in the context of a Black–Scholes model the call price is given by

$$C_t(T,K) = S_t \mathrm{BS}\left((T-t)\sigma^2, \frac{Ke^{-r(T-t)}}{S_t}\right).$$

The quantity $\frac{Ke^{-r(T-t)}}{S_t}$ appearing in the above formula is often called the *moneyness* of the option.

Now notice that $v \mapsto \mathrm{BS}(v,m)$ is strictly increasing and continuous since

$$\frac{\partial \mathrm{BS}}{\partial v}(t,m) = \frac{1}{2\sqrt{v}}\phi\left(-\frac{\log m}{\sqrt{v}} + \frac{\sqrt{v}}{2}\right) > 0$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ is the standard normal density. Hence for fixed m, the function $v \mapsto \mathrm{BS}(v,m)$ can be inverted.

So, given the market price $C_t(T,K)$ of the option, we can find a number $\Sigma_t(T,K)$, called the *implied volatility* of the option, such that

$$C_t(T,K) = S_t \mathrm{BS}((T-t)\Sigma_t(T,K)^2, Ke^{-r(T-t)}/S_t).$$

If the market was still pricing call options by Black–Scholes formula, then there would exist one parameter σ such that $\Sigma_t(T,K) = \sigma$ for all $0 \le t < T$ and $K > 0$. However, in real markets, is is usually the case that the implied volatility surface $(T,K) \mapsto \Sigma_t(T,K)$ is not flat.

One could either conclude Black–Scholes model is the true model of the stock price and that the the market is mispricing options, or that the Black–Scholes model does not quite match reality. The second approach is more prudent. Then, why even consider implied volatility? As Rebonato famously put it:

> Implied volatility is the wrong number to put into wrong formula to obtain the correct price.

However, thanks to the enormous influence of the Black–Scholes theory, the implied volatility is now used as a common language to quote option prices.

8. Local and stochastic volatility models

In the section 7, we have considered the Black–Scholes model–a two asset market model in which the risky asset price is a geometric Brownian motion. The Black–Scholes formula gives an explicit representation of the prices $C_t(T,K)$ of call options in this model in terms of the calendar time t, the current stock price S_t, spot interest rate r, the option maturity T and strike K, and a volatility parameter σ.

However, since the implied volatility surface $\Sigma_t(T, K)$ of real-world option prices is usually not flat, practitioners and researchers have proposed various generalizations of the Black–Scholes model to better match the observed implied volatility surface. We now consider another Markovian model which can match a given implied volatility surface *exactly*.

We consider a model given by

$$\begin{aligned} dB_t &= B_t \, r \, dt \\ dS_t &= S_t(\mu \, dt + \sigma(t, S_t) dW_t). \end{aligned}$$

That is, the idea is replace the constant volatility parameter in Black–Scholes model with a *local volatility* function $\sigma : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$. We will assume that σ is smooth and bounded from below and above. As always, let $\mathbb{Q}$ be the equivalent martingale measures with density process defined by $dZ_t = -Z_t \lambda_t dW_t$ where $\lambda_t = (\mu - r)/\sigma(t, S_t)$.

The next theorem in the present context is usually attributed to Dupire's 1994 paper. However, the existence of a Markovian martingale with a given marginal distribution was proven in 1972 by Kellerer.

THEOREM. *Suppose that*

$$C_0(T, K) = \mathbb{E}^{\mathbb{Q}}[e^{-rT}(S_T - K)^+]$$

Then

$$\frac{\partial C_0}{\partial T}(T, K) = -rK \frac{\partial C_0}{\partial K}(T, K) + \frac{\sigma(T, K)^2}{2} K^2 \frac{\partial^2 C_0}{\partial K^2}(T, K).$$

REMARK. The point of the above theorem is this: Suppose that today's call price surface $\{C_0(T, K) : T > 0, K > 0\}$ is observed from the market. If one chooses the local volatility function σ by *Dupire's formula*

$$\boxed{\sigma(T, K) = \left(\frac{2[\frac{\partial C_0}{\partial T}(T, K) + rK \frac{\partial C_0}{\partial K}(T, K)]}{K^2 \frac{\partial^2 C_0}{\partial K^2}(T, K)} \right)^{1/2}}$$

then the no-abitrage prices of call options in this model exactly match the observed prices.

Of course, if $C_0(T, K) = S_0 \text{BS}(\sigma_0^2 T, Ke^{-rT}/S_0)$ for some constant σ_0 then

$$\left(\frac{2[\frac{\partial C_0}{\partial T}(T, K) + rK \frac{\partial C_0}{\partial K}(T, K)]}{K^2 \frac{\partial^2 C_0}{\partial K^2}(T, K)} \right)^{1/2} = \sigma_0.$$

In general, however, the local volatility surface need not be flat.

PROOF. It can be shown that for all $t > 0$, the random variable S_t has a continuous density with respect to Lebesgue measure; that is, there exists a continuous function $f_{S_t} : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\mathbb{Q}(S_t \le x) = \int_0^x f_{S_t}(y) \, dy.$$

Hence,

$$C_0(T, K) = e^{-rT} \int_0^\infty f_{S_T}(y)(y - K)^+ dy = e^{-rT} \int_K^\infty f_{S_T}(y)(y - K) dy.$$

The fundamental theorem of calculus then implies

$$\begin{aligned}\frac{\partial C_0}{\partial K}(T,K) &= -e^{-rT}\int_K^\infty f_{S_T}(y)\ dy \\ \frac{\partial^2 C_0}{\partial K^2}(T,K) &= e^{-rT} f_{S_T}(K)\end{aligned}$$

[The second equality shows that the risk-neutral density of the stock price at time T can be recovered from the prices of the calls of maturity T. This result was proven by Breeden and Litzenberger in 1978.]

To outline the argument, we proceed formally

$$\begin{aligned}(S_T-K)^+ &= (S_0-K)^+ + \int_0^T \mathbb{1}_{\{S_t\geq K\}} dS_t + \frac{1}{2}\int_0^T \delta_K(S_t) d\langle S\rangle_t \\ &= (S_0-K)^+ + \int_0^T \left(\mathbb{1}_{\{S_t\geq K\}} S_t r + \frac{1}{2}\delta_K(S_t)S_t^2\sigma(t,S_t)^2\right) dt \\ &\quad + \int_0^T \mathbb{1}_{\{S_t\geq K\}} S_t\sigma(t,S_t) d\hat{W}_t\end{aligned}$$

where we have appealed to Itô's formula[4] with $g(x) = (x-K)^+$, $g'(x) = \mathbb{1}_{[K,\infty)}(x)$, and $g''(x) = \delta_K(x)$, the Dirac delta 'function'.

Now computing expected values of both sides

$$(1) \qquad e^{rT}C_0(T,K) = (S_0-K)^+ + \int_0^T\int_K^\infty f_{S_t}(y)y\ r\ dy\ dt + \frac{1}{2}\int_0^T f_{S_t}(K)K^2\sigma(t,K)^2 dt$$

and then differentiating both sides with respect to T yields

$$e^{rT}\left(\frac{\partial C_0}{\partial T}(T,K) + rC_0(T,K)\right) = \int_K^\infty f_{S_T}(y)y\ r\ dy + \frac{1}{2} f_{S_t}(K)K^2\sigma(t,K)^2$$

and the result follows from noting

$$\int_K^\infty f_{S_T}(y)\ y\ dy = \int_0^\infty f_{S_T}(y)(y-K)^+ dy + K\int_K^\infty f_{S_T}(y)dy$$

and applying the appropriate identities. □

Local volatility models are not the end of the story. It turns out that although a local volatility model can be made to exactly match all European call option prices, generally the model fails to correctly price path-dependent options. Therefore, other models have been proposed. These so-called *stochastic volatility* models are of the form

$$\begin{aligned}dB_t &= B_t\ r\ dt \\ dS_t &= S_t(r\ dt + \sigma_t d\hat{W}_t)\end{aligned}$$

where $(\sigma_t)_{t\in\mathbb{R}_+}$ is a given stochastic process. Here are a few popular ones:

- *Local volatility model.* $\sigma_t = \sigma(t,S_t)$
- *CEV model.* $\sigma_t = \gamma S_t^{\beta-1}$
- *Heston model.* $d\sigma_t^2 = \lambda(\bar{\sigma}^2 - \sigma_t^2)dt + \gamma\sigma_t dZ_t$

[4] A version of Itô's formula for non-smooth convex functions, called *Tanaka's formula*, can actually be rigorously stated in terms of a quantity called *local time*.

- *GARCH model.* $d\sigma_t^2 = \lambda(\bar{\sigma}^2 - \sigma_t^2)dt + \gamma\sigma_t^2 dZ_t$
- *SABR model.* $\sigma_t = \gamma_t S_t^{\beta-1}$ and $d\gamma_t = \alpha\gamma_t dZ_t$

where $Z_t = \rho\hat{W}_t + \sqrt{1-\rho^2}W_t^{\perp}$ for an independent Brownian motion $(W_t^{\perp})_{t\in\mathbb{R}_+}$ for $\mathbb{Q}$.

As the economic notion of no-arbitrage is too weak to pin down the precise functional form of a stochastic volatility model, a practioner's choice of model must be made on a combination of issues: how well the model fits data, how easy the model is to calibrate, how quickly a computer can calculate exotic option prices with the model, etc. Notice, however, that aside from the local volatility model (including CEV), stochastic volatility models are incomplete since there more Brownian motions than risky assets.

CHAPTER 5

Interest rate models

1. Bond prices and interest rates

In this last chapter, we explore models for the interest rate term structure. The basic financial instruments in this setting are the zero-coupon bonds.

DEFINITION. A (zero-coupon) *bond* with maturity T is a European contingent claim that pays exactly[1] one unit of currency at time T. We denote by $P(t,T)$ the price at time $t \in [0,T]$ of the bond.

To get a feel for how we should model the bond prices, consider the graph of U.S. Treasury bond prices in Figure 1. Note that on 22 November 2008, the map $T \mapsto P_t(T)$ was decreasing. This is the typical situation. Of course, there are only a finite number of maturities of bonds traded on the the fixed income market. But since this number is very large, it is common practice to represent the zero-coupon bond prices as a continuous curve, rather than a discrete set of points.

We will work in market model where there are bonds of all maturities T available to trade. Therefore, the market model will be specified by a family of processes $\{(P(t,T))_{t\in[0,T]} : T > 0\}$. We also assume that there is a risk-free numéraire process $(B_t)_{t\in\mathbb{R}_+}$, which we will think of as a bank or money market account. Recall that we can write the dynamics of the bank account as

$$dB_t = B_t r_t dt$$

where the adapted process $(r_t)_{t\in\mathbb{R}_+}$ is called the *spot interest rate* or the *short interest rate*. Of course, the above differential equation has the solution

$$B_t = B_0 e^{\int_0^t r_s \; ds}.$$

Now, we formulate a condition so that for any collection of maturities $T_1 < \ldots < T_d$, the market $(B_t, P(t,T_1), \ldots, P(t,T_d))_{t\in[0,T_1]}$ has no arbitrage.

THEOREM. *There is no arbitrage if there exists a locally equivalent measure $\mathbb{Q}$ such that the discounted bond price process $(\tilde{P}(t,T))_{t\in[0,T]}$ is a local martingale for all $T > 0$, where $\tilde{P}(t,T) = P(t,T)/B_t$. In particular, there is no arbitrage if*

$$P(t,T) = \mathbb{E}^{\mathbb{Q}}(e^{-\int_t^T r_s ds} | \mathcal{F}_t)$$

for all $0 \le t \le T$.

[1] We assume that the bond issuer is absolutely credit worthy, and there is zero probability of default. Therefore, we are not discussing corporate bonds, mortgage-backed securities, etc. However, since bonds issued by the U.S. Treasury are backed by the 'full faith and credit' of the U.S. government, they are virtually riskless and will serve as a convenient example.

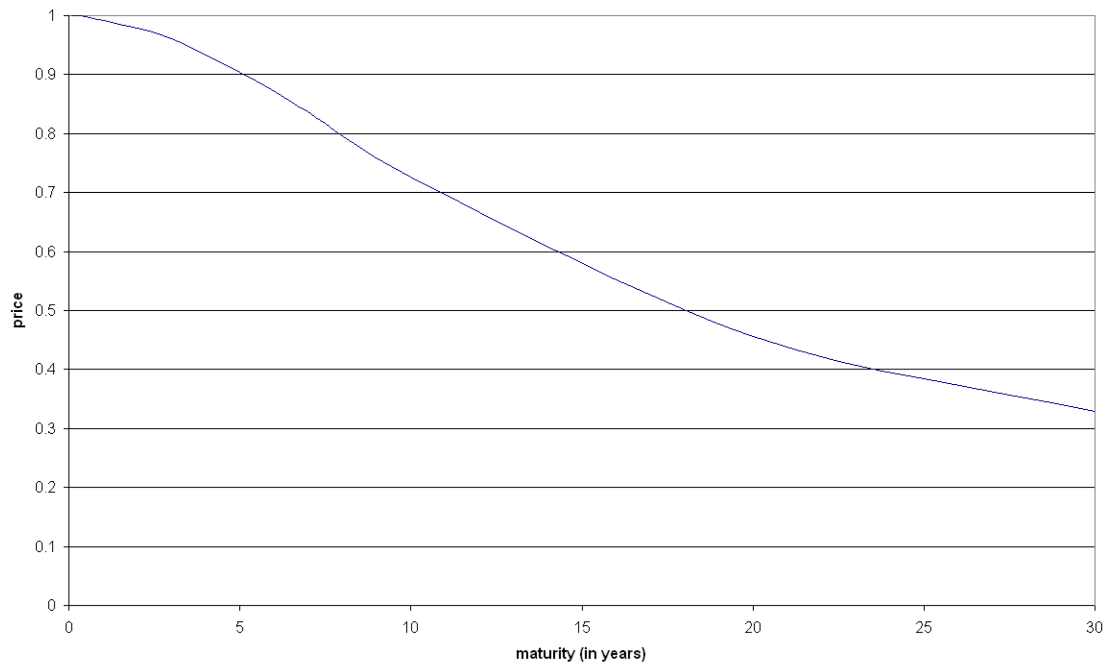


FIGURE 1. Graph of the U.S. Treasury zero-coupon bond price curve on 22 November 2008. Data taken from `http://www.treasury.gov/`

Notice that if

$$P(t,T) = \mathbb{E}^{\mathbb{Q}}(e^{-\int_t^T r_s ds}|\mathcal{F}_t),$$

and r is suitably well behaved, we can differentiate the bond price with respect to maturity to recover the spot rate. Indeed, if the spot rate process is bounded and continuous, then

$$\begin{aligned} -\frac{\partial}{\partial T} P_t(T)\bigg|_{T=t} &= \lim_{T \downarrow t} \mathbb{E}^{\mathbb{Q}} \left(\frac{1 - e^{-\int_t^T r_s ds}}{T-t} |\mathcal{F}_t \right) \\ &= r_t \end{aligned}$$

by the dominated convergence theorem.

From common experience, it seems that we should like to model the interest rate $(r_t)_{t\in\mathbb{R}_+}$ as a non-negative process. Indeed, if $r_t \geq 0$ almost surely for all $t \geq 0$ then the map $T \mapsto P_t(T)$ is decreasing almost surely. However, in actual practice, the interest rate is often modelled by a Gaussian process which might become negative with positive probability.

Rather than speak of bond prices, it is often easier to speak of interest rates. We have already defined the short interest rate. A popular interest rate is the *yield* $y(t,T)$ at time t of a bond maturing at time T defined by the formula

$$y(t,T) = -\frac{1}{T-t} \log P(t,T)$$

Figure 2 is a graph of the U.S. Treasury yield curve on 22 November 2008. The yield curve and the bond price curve contain the same information, since

$$P(t,T) = e^{-(T-t)\ y(t,T)}.$$

Note that spot interest rate is just the left hand point of the yield curve $r_t = y(t,t)$.

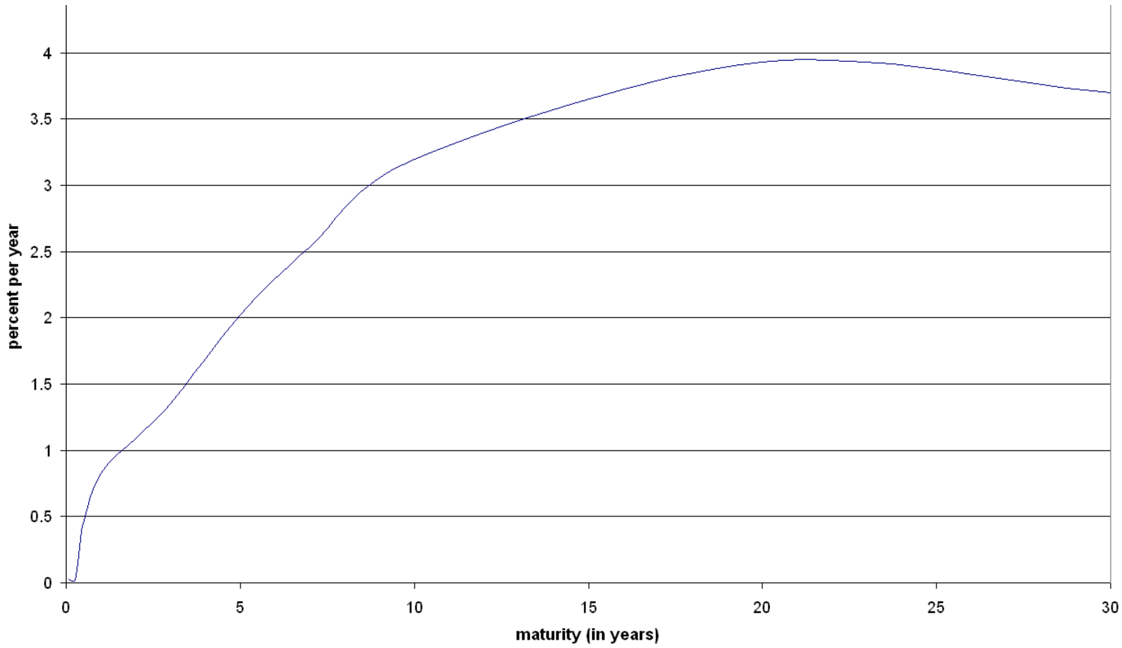


FIGURE 2. Graph of the U.S. Treasury yield curve on 22 November 2008. Data taken from `http://www.treasury.gov/`

For us, a more useful interest rate is the *forward rate* $f(t,T)$ at time t for maturity T, defined by

$$f(t,T) = -\frac{\partial}{\partial T} \log P(t,T).$$

Again, notice that the spot rate is the left hand end point of the forward rate curve $r_t = f(t,t)$.

Note that if $(r_t)_{t\in\mathbb{R}_+}$ is suitably regular, say bounded and continuous, then

$$f(t,T) = \frac{\mathbb{E}^{\mathbb{Q}}(r_T e^{-\int_t^T r_s ds}|\mathcal{F}_t)}{\mathbb{E}^{\mathbb{Q}}(e^{-\int_t^T r_s ds}|\mathcal{F}_t)}$$

so that the forward rate can be interpreted as the $\mathcal{F}_t$-measurable random variable $f(t,T)$ such that the no-arbitrage price at time t of the claim that pays $r_T - f(t,T)$ at time T is zero.

Again, note that the forward rates contain the same information as the bond pricese since

$$P(t,T) = e^{-\int_t^T f(t,s)\ ds}.$$

The *term structure* of interest rates refers the function $T \mapsto P(t,T)$, or equivalently, the price data encoded in either of the functions $T \mapsto y(t,T)$ or $T \mapsto f(t,T)$.

2. Short rate models

We begin with a market that has just the bank account B. We will consider an Itô process short interest rate model of the form

$$dr_t = a_t dt + \beta_t dW_t$$

for adapted process $(a_t)_{t\mathbb{R}_+}$ and $(b_t)_{t\in\mathbb{R}_+}$, and a Brownian motion $(W_t)_{t\in\mathbb{R}_+}$ for $\mathbb{P}$.

Note that while in a complete stock market model there was only one way to switch to an equivalent martingale measure, no such choice is possible since the short rate is not traded. However, we know that there is no arbitrage if the market somehow picks an equivalent martingale measure $\mathbb{Q}$ to price the bonds. We will assume that the market price of risk is given by the process $(\lambda_t)_{t\in\mathbb{R}_+}$ so that

$$dr_t = \alpha_t dt + \beta_t d\hat{W}_t$$

where $d\hat{W}_t = dW_t + \lambda_t\, dt$ defines a Brownian motion for the measure $\mathbb{Q}$ whose density process is given by $dZ_t = -Z_t\lambda_t dW_t$, and where $\alpha_t = a_t - \beta_t\lambda_t$ defines the risk-neutral drift.

Since we are interested in pricing and hedging, there is no need to model the processes $(a_t)_{t\in\mathbb{R}_+}$ and $(\lambda_t)_{t\in\mathbb{R}_+}$ separately. However, we must be careful to realize that is impossible to estimate the distribution of the random variable α_t directly from a time series $r_{t_1}, \ldots, r_{t_n}$.

We now study the case when the short rate is Markovian. Assume that

$$dr_t = \alpha(t, r_t)dt + \beta(t, r_t)d\hat{W}_t$$

for some non-random functions $\alpha : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ and $\beta : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$.

As we have learned for Markovian stock models, the price of contingent claims can be expressed in terms the solution of a PDE:

THEOREM. *Fix $T > 0$ and suppose $V : [0, T] \times \mathbb{R} \to \mathbb{R}$ satisfies the PDE*

$$\begin{aligned}
&\frac{\partial V}{\partial t}(t, r) + \alpha(t, r)\frac{\partial V}{\partial r}(t, r) + \frac{1}{2}\beta(t, r)^2\frac{\partial^2 V}{\partial r^2}(t, r) = rV(t, r)\\
&V(T, r) = 1
\end{aligned}$$

If

$$P(t, T) = V(t, r_t)$$

then there is no arbitrage.

PROOF. Itô's formula implies

$$\begin{aligned}
d\left(e^{-\int_0^t r_s ds}V(t, r_t)\right) &= -r_t e^{-\int_0^t r_s ds}V(t, r_t)dt\\
&\quad + e^{-\int_0^t r_s ds}\left(\frac{\partial V}{\partial t}(t, r_t)\, dt + \frac{\partial V}{\partial r}(t, r_t)dr_t + \frac{1}{2}\frac{\partial^2 V}{\partial r^2}(t, r_t)d\langle r\rangle_t\right)\\
&= e^{-\int_0^t r_s ds}\Big(\frac{\partial V}{\partial t}(t, r_t) + \alpha(t, r_t)\frac{\partial V}{\partial r}(t, r_t) + \frac{1}{2}\beta(t, r_t)^2\frac{\partial^2 V}{\partial r^2}(t, r_t)\\
&\quad - r_t V(t, r_t)\Big)dt + e^{-\int_0^t r_s ds}\frac{\partial V}{\partial r}(t, r_t)\beta(t, r_t)d\hat{W}_t
\end{aligned}$$

Since the drift vanishes by assumption, so $(\tilde{P}(t, T))_{t\in[0,T]}$ is a local martingale. ∎

2.1. Vasicek model. In 1977, Vasicek proposed the following model for the short rate:

$$dr_t = \lambda(\bar{r} - r_t)dt + \sigma d\hat{W}_t$$

for a parameter $\bar{r} > 0$ interpreted as a mean short rate, a mean-reversion parameter $\lambda > 0$, and a volatility parameter $\sigma > 0$. This stochastic differential equation can be solved explicitly to yield

$$r_t = e^{-\lambda t} r_0 + (1 - e^{-\lambda t})\bar{r} + \int_0^t e^{-\lambda(t-s)} \sigma d\hat{W}_s.$$

Note that for each $t \geq 0$ the random variable r_t is Gaussian[2] under the measure $\mathbb{Q}$ with

$$\mathbb{E}^{\mathbb{Q}}(r_t) = e^{-\lambda t} r_0 + (1 - e^{-\lambda t})\bar{r} \quad \text{and} \quad \mathrm{Var}^{\mathbb{Q}}(r_t) = \int_0^t e^{-2\lambda(t-s)} \sigma^2 ds = \frac{\sigma^2}{2\lambda}(1 - e^{-2\lambda t}).$$

Moreover, one can show that the process is ergodic and converges to the invariant distribution $N\left(\bar{r}, \frac{\sigma^2}{2\lambda}\right)$. In particular, we have

$$\frac{1}{T}\int_0^T r_s\, ds \to \bar{r} \quad \mathbb{Q} - \text{almost surely}.$$

Please note, however, that in the present framework we can say *absolutely nothing* about the distribution of r_t for the objective measure $\mathbb{P}$, unless we have a model for the market price of risk.

Since the short rate r_t is Gaussian, the advantage of this type of model is that it is relatively easy to compute prices, for instance of bonds, explicitly. A disadvantage of this model is that there is a chance that $r_t < 0$ for some time $t > 0$. Recall that a normal random variable can take any real value, both positive and negative. However, for sensible parameter values, the $\mathbb{Q}$-probabilty of the event $\{r_t < 0\}$ is pretty small.

We can also use the above theorem to compute bond prices. Indeed, fix $T > 0$ and consider the PDE

$$\begin{aligned}
&\frac{\partial V}{\partial t}(t,r) + \lambda(\bar{r} - r)\frac{\partial V}{\partial r}(t,r) + \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial r^2}(t,r) = rV(t,r) \\
&V(T,r) = 1.
\end{aligned}$$

We can make the ansatz

$$V(t,r) = e^{-rA(t) - B(t)}$$

for some functions A and B which satisfy the boundary conditions $A(T) = B(T) = 0$. Substituting this into the PDE yields

$$(-A'(t)r - B'(t))V(t,r) - \lambda(\bar{r} - r)A(t)V(t,r) + \frac{\sigma^2}{2}A(t)^2 V(t,r) = rV(t,r).$$

Since this is supposed to be an identity for all (t, r) we have

$$\begin{aligned}
A'(t) &= \lambda A(t) - 1 \\
B'(t) &= -\lambda \bar{r} A(t) + \frac{\sigma^2}{2}A(t)^2
\end{aligned}$$

[2] A continuous Gaussian Markov process is often called a *Ornstein–Uhlenbeck* process.

The solution to this pair of equations is

$$\begin{aligned} A(t,T) &= \frac{1}{\lambda}(1-e^{-\lambda(T-t)}) \\ B(t,T) &= \int_t^T \left(\lambda\bar{r}A(s) - \frac{\sigma^2}{2}A(s)^2\right) ds \end{aligned}$$

so that the bond price is given by

$$P(t,T) = \exp\left(-r_t\frac{(1-e^{-\lambda(T-t)})}{\lambda} - \bar{r}\int_0^{T-t}(1-e^{-\lambda u})\, du + \frac{\sigma^2}{2\lambda^2}\int_0^{T-t}(1-e^{-\lambda u})^2\, du\right).$$

This is a mess. However, the forward rates are more manageable:

$$\boxed{f(t,t+x) = r_t e^{-\lambda x} + \bar{r}(1-e^{-\lambda x}) - \frac{\sigma^2}{2\lambda^2}(1-e^{-\lambda x})^2}$$

This formula says that for the Vasicek model, the forward rates at time t are an affine function of the short rate at time t. (An affine function is of the form $g(x) = ax+b$, that is, its graph is a line.)

2.2. Cox–Ingersoll-Ross model. In 1985, Cox, Ingersoll, and Ross proposed the following model for the short rate:

$$dr_t = \lambda(\bar{r} - r_t) + \sigma\sqrt{r_t}d\hat{W}_t$$

for a parameter $\bar{r} > 0$ interpreted as a mean short rate, a mean-reversion parameter $\lambda > 0$, and a volatility parameter $\sigma > 0$. The process $(r_t)_{t\in\mathbb{R}_+}$ satisfying the above stochastic differential equation is often called a square-root diffusion or CIR process, though this stochastic process was studied as early as 1951 by Feller. Althought the stochastic differential equation cannot be solved explicitly, one can say quite a lot about this process. For instance, one can show that the process is ergodic and its invariant disribution is a gamma distribution with mean $\bar{r}$.

An advantage of this model over the Vasicek model is that the short rate r_t is non-negative for all $t \geq 0$. Furthermore, explicit formula are still available for the bond prices. Indeed, the CIR model is another example of an affine term structure model: Again, fix $T > 0$ and consider the PDE

$$\begin{aligned} &\frac{\partial V}{\partial t}(t,r) + \lambda(\bar{r}-r)\frac{\partial V}{\partial r}(t,r) + \frac{1}{2}\sigma^2 r\frac{\partial^2 V}{\partial r^2}(t,r) = rV(t,r) \\ &V(T,r) = 1. \end{aligned}$$

As before we can make the ansatz

$$V(t,r) = e^{-rA(t)-B(t)}$$

for some functions A and B which satisfy the boundary conditions $A(T) = B(T) = 0$. Substituting this into the PDE yields

$$(-A'(t)r - B'(t))V(t,r) - \lambda(\bar{r}-r)A(t)V(t,r) + \frac{\sigma^2}{2}rA(t)^2V(t,r) = rV(t,r).$$

This time we have

$$\begin{aligned} A'(t) &= \lambda A(t) + \frac{\sigma^2}{2} A(t)^2 - 1 \\ B'(t) &= -\lambda \bar{r} A(t). \end{aligned}$$

The equation for A is a Riccati equation, whose solution is

$$\begin{aligned} A(t,T) &= \frac{2(e^{\gamma(T-t)} - 1)}{(\gamma+\lambda)e^{\gamma(T-t)} + (\gamma - \lambda)} \\ B(t,T) &= \int_t^T \lambda \bar{r} A(s) ds \end{aligned}$$

where $\gamma = \sqrt{\lambda^2 + 2\sigma^2}$. The bond prices are too messy to write down, but the forward rates are given by

$$\boxed{f(t,t+x) = \frac{4\gamma^2 e^{\gamma x}}{[(\gamma+\lambda)e^{\gamma x} + (\gamma-\lambda)]^2} r_t + \frac{2\lambda\bar{r}(e^{\gamma x} - 1)}{(\gamma+\lambda)e^{\gamma x} + (\gamma - \lambda)}.}$$

In particular, the forward rates for the CIR model are again given by an affine function of the short rate.

3. Factor models

The Markovian short rate models are popular in practice, especially the Vasicek and CIR models in which formulas for the bond prices are available in closed form.

However, a possible shortcoming of these models is that they predict a very rigid term structure. Indeed, there is very little flexibility in the possible shapes of the forward rate curve, since there exists a deterministic function $g : D \times \mathbb{R} \to \mathbb{R}$ such that

$$f(t,T) = g(t,T,r_t)$$

where $D = \{(t,T) : 0 \le t \le T\}$. In particular, in the Vasicek and CIR models, the function $r \mapsto g(t,T,r)$ is affine, so that the correlation coefficient

$$\rho(r_t, f(t,T)) = 1$$

for all $0 < t \le T$. In this section, we consider more general factor models, of which the short rate models are only a special case.

The idea is to assume that there are d underlying economic 'factors' in the market. We model these factors as the solution $(Z_t)_{t\in\mathbb{R}_+}$ of a stochastic differential equation

$$dZ_t = a(t, Z_t)dt + b(t, Z_t)d\hat{W}_t$$

where $a : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ and $b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are given functions and $(\hat{W}_t)_{t\in\mathbb{R}_+}$ is a d-dimensional Brownian motion for the equivalent martingale measure $\mathbb{Q}$. We then assume that the short rate is given by a function $r_t = R(t, Z_t)$. The short rate models considered in the last section have $d = 1$, $Z_t = r_t$, and $R(t,r) = r$.

The main theorem is below.

THEOREM. *Fix $T > 0$. Suppose $V : [0,T] \times \mathbb{R}^d \to \mathbb{R}_+$ satifies the partial differential equation*

$$\frac{\partial V}{\partial t} + \sum_{i=1}^{d} a_i \frac{\partial V}{\partial Z_i} + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} B_{ij} \frac{\partial^2 V}{\partial Z_i \partial Z_j} = RV$$
$$V(T, Z) = 1$$

where $B(t,z) = b(t,z)b(t,z))^{\mathrm{T}}$. If $P(t,T) = V(t, Z_t)$ then the market consisting of the bank account and bond maturity T has no arbitrage.

We now consider a special case of the above theorem, first studied by Duffie and Kan in 1996. We let

$$dZ_t = (\alpha + \beta Z_t)dt + \begin{pmatrix} \sqrt{\gamma_1 + \delta_1 \cdot Z_t} & 0 & \cdots & 0 \\ 0 & \sqrt{\gamma_2 + \delta_2 \cdot Z_t} & \cdots & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \sqrt{\gamma_d + \delta_d \cdot Z_t} \end{pmatrix} d\hat{W}_t,$$

Here, $\gamma_1, \ldots \gamma_d$ are d positive real constants, α and $\delta_1, \ldots, \delta_d$ are d-dimensional constant vectors, and β is a $d \times d$ constant matrix.

Suppose the short rate is given by

$$r_t = Z_t^{(1)} + \ldots + Z_t^{(d)}.$$

REMARK. The analysis of such a stochastic differential equation is very simple when all the δ_i's are zero. In this case, there is existence and uniqueness of a solution, and in fact, the solution is a d-dimensional Ornstein–Uhlenbeck process. However, the situation is much more delicate when some of the δ_i's are non-zero. Existence and uniqueness of solutions of such stochastic differential equations is not guaranteed, and analyzing the properties of this non-Gaussian diffusion is not easy because of the random volatility created by the non-zero δ_i's.

As in the case of the Vasicek and CIR models, we can make the ansatz that the bond prices can be written in the exponential affine form

$$P(t,T) = e^{-A(t,T)\cdot Z_t - B(t,T)}.$$

for deterministic functions $A : D \to \mathbb{R} \to \mathbb{R}^d$ and $B : D \to \mathbb{R}$.

As before, the functions A and B can be found by solving the following system of $d+1$ coupled Riccati equations

$$\begin{aligned} \frac{\partial A_k}{\partial t} &= -\sum_{i=1}^{d} \beta_{ik} A_i + \frac{1}{2} \delta_k A_k^2 - 1 \\ \frac{\partial B}{\partial t} &= -\sum_{i=1}^{d} \alpha_i A_i + \frac{1}{2} \sum_{i=1}^{d} \gamma_i A_i^2. \end{aligned}$$

These equations can be solved numerically, for instance, subject to the boundary conditions

$$A_k(T,T) = 0 = B(T,T)$$

for all $k \in \{1, \dots, d\}$.

Given the exponential affine bond prices, note that

$$f(t,T) = \frac{\partial A}{\partial T}(t,T) \cdot Z_t + \frac{\partial B}{\partial T}(t,T).$$

Fix d dates $T_1, \dots, T_d$ and consider the d benchmark rates $f(t,T_1), \dots, f(t,T_d)$. If the matrix

$$\Gamma_t = \left(\frac{\partial A_j}{\partial T}(t,T_i) \right)_{i,j \in \{1,\dots,d\}}$$

is invertible, then we can recover the factors as linear combinations of the benchmark forward rates:

$$Z_t = \Gamma_t^{-1} \left[\left(f(t,T_i) - \frac{\partial B}{\partial T}(t,T_i) \right)_{i \in \{1,\dots,d\}} \right]$$

4. The Heath–Jarrow–Morton framework

Starting from a factor model, the derived bond prices are necessarily Itô processes. There is no arbitrage in a factor model since, by construction, there exists an equivalent martingale measure $\mathbb{Q}$ such that all discounted bond prices $(\tilde{P}(t,T))_{t \in [0,T]}$ are local martingales, where the discounted bond price at time t for maturity T is given by

$$\tilde{P}(t,T) = e^{-\int_0^t r_s ds} P_t(T).$$

The insight of Heath, Jarrow, and Morton in 1992 was that we can change perspectives by modelling the bond prices directly.

Motivation. Indeed, suppose we start out with just the bond market, but without the bank account. We can construct the bank account by considering an investor holding his wealth in just-maturing bonds. More concretely, suppose at time 0 the investor has B_0 units of wealth. Fix a sequence $0 \le t_0 < t_1 < \dots$ of times and suppose that during the interval $(t_{i-1}, t_i]$ the investor holds all of his wealth in the bond which matures at time t_i. If the investor's wealth at time t is denoted by B_t, and the number of shares of the just-maturing bond by π_t, the budget constraint is

$$B_{t_{i-1}} = \pi_{t_i} P(t_{i-1}, t_i)$$

and the self-financing condition is

$$B_{t_i} = \pi_{t_i}$$

since $P(t,t) = 1$ for all t. Hence, the rate of change of the wealth is given by

$$\frac{B_{t_i} - B_{t_{i-1}}}{t_i - t_{i-1}} = \frac{B_{t_{i-1}}}{P_{t_{i-1}}(t_i)} \frac{1 - P(t_{i-1}, t_i)}{t_i - t_{i-1}}$$

By taking the limit as $t_i - t_{i-1} \to 0$, we can *define* the spot rate by

$$r_t = -\frac{\partial}{\partial T} P(t,T)|_{T=t}$$

so that $dB_t = B_t r_t dt$ as before.

The usual formulation of the HJM idea is in terms of the forward rates. As usual, we put ourselves in the context of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which we can define a d-dimensional Brownian motion $(W_t)_{t\in\mathbb{R}_+}$.

THEOREM. *Suppose for each T, the foward rate process $(f(t,T))_{t\in[0,T]}$ has dynamics*

$$df(t,T) = a(t,T)dt + \sum_{i=1}^{n} \sigma^{(i)}(t,T)dW_t^{(i)}$$

for some suitably regular adapted processes $(a(t,T))_{t\in[0,T]}$ and $(\sigma^{(i)}(t,T))_{t\in[0,T]}$. Let the short rate be given by $r_t = f(t,t)$ and the bank account dynamics by

$$dB_t = B_t \; r_t \; dt.$$

Finally, let the bond prices be given by

$$P(t,T) = e^{-\int_t^T f(t,s) \; ds}.$$

If there exists a d-dimensional bounded adapted process $(\lambda_t)_{t\in\mathbb{R}_+}$ such that

$$a(t,(T) = \sum_{i=1}^{d} \sigma^{(i)}(t,T)\left(\lambda_t^{(i)} + \int_t^T \sigma^{(i)}(t,s)ds\right),$$

then, for any set of d maturities $0 < T_1 < \ldots < T_d$, the market model with prices $(B_t, P(t,T_1), \ldots, P(t,T_d))_{t\in}$ has no arbitrage.

REMARK. The upshot of the HJM result is that the drift and the volatilty of the forward rate dynamics cannot be prescribed independently. Indeed, they must be related by the famous formula

$$\boxed{a(t,T) = \sigma(t,T)\cdot\left(\lambda_t + \int_t^T \sigma(t,s)ds\right),}$$

usually called the HJM drift condition. Notice that this drift/volatility contraint is not present in models in which only the dynamics of the short rate are specified, as in Section 2.

The difference with the short rate models is that we are now trying to model the dynamics of the whole term structure. Indeed, in the HJM framework, we can initialize the model with *any* initial forward rate curve $T \mapsto f(0,T)$. Nevertheless, note that any of the short rate or factor models can be put into the HJM framework, just by choosing the initial forward rate curve to match the one predicted by the model.

PROOF. Define a locally equivalent measure $\mathbb{Q}$ by the density process $dZ_t = -Z_t\lambda_t \cdot dW_t$. For this measure, the process defined by $d\hat{W}_t = dW_t + \lambda_t \; dt$ is a Brownian motion. We can rewrite the forward rate dynamics as

$$df(t,T) = \sigma(t,T)\cdot\int_t^T \sigma(t,s)ds \; dt + \sigma(t,T)\cdot d\hat{W}_t.$$

It is enough to show that for each $T > 0$, the discounted bond price process $\tilde{P}(t,T) = e^{-\int_0^t r_s ds - \int_t^T f(t,s)ds}$ is a local martingale. Now applying some formal manipulations[3]

$$d\left(\int_0^t r_s\ ds + \int_t^T f(t,s)\ ds\right) = (r_t - f(t,t))dt + \int_t^T df(t,s)\ ds$$
$$= \frac{1}{2}\left|\int_t^T \sigma(t,s)ds\right|^2 dt + \int_t^T \sigma(t,s)ds \cdot d\hat{W}_t.$$

Hence, by Itô's formula, we have

$$d\tilde{P}(t,T) = -\tilde{P}(t,T)\int_t^T \sigma(t,s)ds \cdot d\hat{W}_t$$

and we're done. □

We conclude this section with some examples. In these examples, the forward rates are Gaussian under the measure $\mathbb{Q}$, and hence are vulnerable to the criticism that there is a positive probability that the rates become negative.

---

[3]We would like to appeal to a *stochastic Fubini theorem* in order to exchange the order of integration in the double integral. The question is: if we fix $S, T \geq 0$, when do we have the equality

$$\int_0^T \int_0^S g(s,t)ds\ dW_t = \int_0^S \int_0^T g(s,t)dW_t\ ds?$$

First note that the equality holds for $g \in \mathcal{S}$ where $\mathcal{S}$ is the set

$$\mathcal{S} = \left\{\sum_{i=1}^n k_i \mathbb{1}_{(s_{i-1},s_j]\times(t_{i-1},t_i]} : 0 \leq s_{i-1} < s_i \leq S, 0 \leq t_{i-1} < t_i \leq T, k_i \text{ is bounded and} \mathcal{F}_{t_{i-1}} \text{ measurable}\right\}.$$

Now suppose there exists a sequence $(g_n)_{n\in\mathbb{N}}$ in $\mathcal{S}$ such that

$$\mathbb{E}\left(\int_0^T \int_0^S [g(s,t) - g_n(s,t)]^2 ds\ dt\right) = \mathbb{E}\left(\int_0^S \int_0^T [g(s,t) - g_n(s,t)]^2 dt\ ds\right) \to 0.$$

Then g satisfies the exchange of order of integration equality, since

$$\mathbb{E}\left[\left(\int_0^T \int_0^S g(s,t)ds\ dW_t - \int_0^T \int_0^S g_n(s,t)ds\ dW_t\right)^2\right] = \mathbb{E}\left[\int_0^T \left(\int_0^S (g(s,t) - g_n(s,t))ds\right)^2 dt\right]$$
$$= S\ \mathbb{E}\left[\int_0^T \int_0^S (g(s,t) - g_n(s,t))^2 ds\ dt\right] \to 0$$

by Itô's isometry and the Cauchy–Schwarz inequality, and

$$\mathbb{E}\left[\left(\int_0^S \int_0^T g(s,t)dW_t\ ds - \int_0^T \int_0^T g_n(s,t)dW_t\ ds\right)^2\right] \leq S\ \mathbb{E}\left[\int_0^S \left(\int_0^T (g(s,t) - g_n(s,t))dW_t\right)^2 dt\right]$$
$$= S\ \mathbb{E}\left[\int_0^S \int_0^T (g(s,t) - g_n(s,t))^2 dt\ ds\right] \to 0$$

by the Cauchy–Schwarz inequality, Fubini's theorem, and Itô's isometry. Finally, we can replace the condition $\mathbb{E}\left(\int_0^T \int_0^S g(s,t)^2 ds\ dt\right) < \infty$ with $\int_0^T \int_0^S g(s,t)^2 ds\ dt < \infty$ a.s. by localization.

An example which will occur frequently is when g is not random and continuous (or at least Riemann integrable) on $[0,S] \times [0,T]$.

4.1. Ho–Lee. (1986) This model is the simplest possible model HJM model. Let $d = 1$ and $\sigma(t,T) = \sigma_0$ be constant. Then

$$df(t,T) = \sigma_0^2(T-t)\ dt + \sigma_0 d\hat{W}_t.$$

or

$$f(t,T) = f(0,T) + \sigma_0^2(Tt - t^2/2) + \sigma_0 \hat{W}_t.$$

Here is an unusual feature of this model: if the initial forward rate curve $T \mapsto f(0,T)$ is bounded from below, then for positive times t the forward rates $f(0,T) \to \infty$ as $T \to \infty$.

The short rate is then given by

$$r_t = f(0,t) + \sigma_0^2 t^2/2 + \sigma_0 \hat{W}_t.$$

Hence the Ho–Lee model corresponds to the following short rate model:

$$dr_t = (f_0'(t) + \sigma_0^2 t)dt + \sigma_0 d\hat{W}_t.$$

4.2. Vasicek–Hull–White. (1990) Again let $d = 1$ but now $\sigma(t,T) = \sigma_0 e^{-\lambda(T-t)}$ for positive constants σ_0 and λ. Then

$$df(t,T) = \frac{\sigma_0^2}{\lambda} e^{-\lambda(T-t)}(1 - e^{-\lambda(T-t)})dt + \sigma_0 e^{-\lambda(T-t)} d\hat{W}_t.$$

The short rates are given by

$$\begin{aligned} r_t &= f(0,t) + \int_0^t \frac{\sigma_0^2}{\lambda} e^{-\lambda(t-s)}(1 - e^{-\lambda(t-s)})ds + \int_0^t \sigma_0 e^{-\lambda(t-s)} d\hat{W}_s \\ &= f(0,t) + \frac{\sigma_0^2}{2\lambda^2}(1 - e^{-\lambda t})^2 + \int_0^t \sigma_0 e^{-\lambda(t-s)} d\hat{W}_s \end{aligned}$$

The short rate dynamics are given by

$$\begin{aligned} dr_t &= \left(f_0'(t) + \frac{\sigma_0^2}{\lambda} e^{-\lambda t}(1 - e^{-\lambda t}) \right) dt + \sigma_0 d\hat{W}_t - \lambda \int_0^t \sigma_0 e^{-\lambda(t-s)} d\hat{W}_s\ dt \\ &= \left(f_0'(t) + \lambda f_0(t) + \frac{\sigma_0^2}{2\lambda}(1 - e^{-2\lambda t}) - \lambda r_t \right) dt + \sigma_0 d\hat{W}_t \end{aligned}$$

Hence, the Hull–White extension of the Vasicek essentially replaces the mean interest rate $\bar{r}$ with a time-varying, but non-random, mean rate $\bar{r}(t)$.

4.3. Kennedy. (1994) Note that for the HJM models discussed above, the forward rates are given by

$$f(t,T) = f(0,T) + \int_0^t \sigma(u,T) \cdot \int_u^T \sigma(u,s)ds\ du + \int_0^t \sigma(u,T) \cdot d\hat{W}_u.$$

If σ is not random, then the distribution of $f(t,T)$ under the risk-neutral measure $\mathbb{Q}$ is Gaussian with mean

$$\mathbb{E}^{\mathbb{Q}}[f(t,T)] = f_0(T) + \int_0^t \sigma(u,T) \cdot \int_u^T \sigma(u,s)ds\ du$$

and covariance

$$\mathrm{Cov}^{\mathbb{Q}}[f(s,S), f(s,T)] = \int_0^{s \wedge t} \sigma(u,S) \cdot \sigma(u,T)du.$$

Kennedy reversed this logic, and considered a Gaussian random field $\{f(t,T) : 0 \le t \le T\}$ with mean $\mu(t,T)$ and covariance $C(s,t;S,T)$. Suppose that covariance has the special form

$$C(s,t;S,T) = c_{s\wedge t}(S,T)$$

so that, for each fixed $T > 0$, the increments of $(f(t,T))_{t\in[0,T]}$ are independent. Then there is no arbitrage if the mean is given by

$$\mu(t,T) = f(0,T) + \int_0^T c_{t\wedge s}(s,T)ds.$$

An advantage of this formulation of the Gaussian HJM model is that one is no longer restricted to finite dimensional Brownian motions, and, therefore, there is much more flexibility to specify the correlation of the increments. For instance, one choice is to have the correlation of the increments decay exponentially in the difference of the maturities:

$$`\rho(df_t(S), df_t(T)) = e^{-\beta|T-S|}.'$$

This can be achieved by taking c to be

$$c_t(S,T) = e^{-\beta|T-S|}\int_0^t \alpha_u(S)\alpha_u(T)du$$

for real valued functions α_u.

CHAPTER 6

Crashcourse on probability theory

These notes are a list of many of the definitions and results of probability theory needed to follow the Advanced Financial Models course. Since they are free from any motivating exposition or examples, and since no proofs are given for any of the theorems, these notes should be used only as a reference. A table of notation is in the appendix.

1. Measures

DEFINITION. Let Ω be a set. A *sigma-field* on Ω is a non-empty set $\mathcal{F}$ of subsets of Ω such that

(1) if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$,
(2) if $A_1, A_2, \ldots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

The terms sigma-field and *sigma-algebra* are interchangeable.

The *Borel* sigma-field $\mathcal{B}$ on $\mathbb{R}$ is the smallest sigma-field containing every open interval. More generally, if Ω is a topological space, for instance $\mathbb{R}^n$, the Borel sigma-field on Ω is the smallest sigma-field containing every open set.

DEFINITION. Let Ω be a set and let $\mathcal{F}$ be a sigma-field on Ω. A *measure* μ on the *measurable space* $(\Omega, \mathcal{F})$ is a $\mu : \mathcal{F} \to [0, \infty]$ such that

(1) $\mu(\emptyset) = 0$
(2) if $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint then $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.

THEOREM. *There exists a unique measure* Leb *on* $(\mathbb{R}, \mathcal{B})$ *such that*

$$\text{Leb}(a, b] = b - a$$

for every $b > a$. *This measure is called* Lebesgue *measure.*

DEFINITION. A *probability measure* $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is a measure such that $\mathbb{P}(\Omega) = 1$.

Let Ω be a set, $\mathcal{F}$ a sigma-field on Ω, and $\mathbb{P}$ a probability measure on $(\Omega, \mathcal{F})$. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space*.

The set Ω is called the *sample space*, and an element of Ω is called an *outcome*. A subset of Ω which is an element of $\mathcal{F}$ is called an *event*.

Let $A \in \mathcal{F}$ be an event. If $\mathbb{P}(A) = 1$ then A is called an *almost sure* event, and if $\mathbb{P}(A) = 0$ then A is called a *null* event. The phrase 'almost surely' is often abbreviated *a.s.* A sigma-field is called *trivial* if each of its elements is either almost sure or null.

2. Random variables

DEFINITION. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *random variable* is a function $X : \Omega \to \mathbb{R}$ such that the set $\{\omega \in \Omega : X(\omega) \le t\}$ is an element of $\mathcal{F}$ for all $t \in \mathbb{R}$.

Let A be a subset of $\mathbb{R}$, and let X be a random variable. We use the notation $\{X \in A\}$ to denote the set $\{\omega \in \Omega : X(\omega) \in A\}$. For instance, the event $\{X \leq t\}$ denotes $\{\omega \in \Omega : X(\omega) \leq t\}$.

The *distribution function* of X is the function $F_X : \mathbb{R} \to [0, 1]$ defined by

$$F_X(t) = \mathbb{P}(X \leq t)$$

for all $t \in \mathbb{R}$.

We also use the term random variable to refer to measurable functions X from Ω to more general spaces. In particular, we call a function $X : \Omega \to \mathbb{R}^n$ a random variable or *random vector* if $X(\omega) = (X_1(\omega), \ldots, X_n(\omega))$ and X_i is a random variable for each $i \in \{1, \ldots, n\}$.

Definition. Let A be an event in Ω. The *indicator function* of the event A is the random variable $\mathbb{1}_A : \Omega \to \{0, 1\}$ defined by

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \in A^c \end{cases}$$

for all $\omega \in \Omega$.

3. Expectations and variances

Definition. Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. The *expected value* of X is denoted by $\mathbb{E}(X)$ and is defined as follows

- X is *simple*, i.e. takes only a finite number of values $x_1, \ldots, x_n$.

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i).$$

- $X \geq 0$ almost surely.

$$\mathbb{E}(X) = \sup\{\mathbb{E}(Y) : Y \text{ simple and } 0 \leq Y \leq X \text{a.s.}\}$$

  Note that the expected value of a non-negative random variable may take the value ∞.

- Either $\mathbb{E}(X^+)$ or $\mathbb{E}(X^-)$ is finite.

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$$

- X is vector valued and $\mathbb{E}(|X|) < \infty$.

$$\mathbb{E}[(X_1, \ldots, X_d)] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])$$

A random variable X is *integrable* iff $\mathbb{E}(|X|) < \infty$ and is *square-integrable* iff $\mathbb{E}(X^2) < \infty$. The terms expected value, *expectation*, and *mean* are interchangeable.

The *variance* of an integrable random variable X, written $\mathrm{Var}(X)$, is

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

The *covariance* of square-integrable random variable X and Y, written $\mathrm{Cov}(X, Y)$, is

$$\mathrm{Cov}(X, Y) = \mathbb{E}\{[X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]\} = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

If neither X or Y is almost surely constant, then their correlation, written $\rho(X, Y)$, is

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)^{1/2}\mathrm{Var}(Y)^{1/2}}.$$

Random variables X and Y are called *uncorrelated* if $\operatorname{Cov}(X,Y) = 0$.

THEOREM. *Let X and Y be integrable random variables.*

- *linearity:* $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$
- *positivity: Suppose $X \geq 0$ almost surely. Then $\mathbb{E}(X) \geq 0$ with equality if and only if $X = 0$ almost surely.*

DEFINITION. For $p \geq 1$, the space L^p is the collection of random variables such that $\mathbb{E}(|X|^p) < \infty$. The space L^∞ is the collection of random variables which are bounded almost surely.

THEOREM (Jensen's inequality). *Let X be a random variable and $g : \mathbb{R} \to \mathbb{R}$ be a convex function. Then*

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

whenever the expectations exist. If g is strictly convex, the above inequality is strict unless X is constant.

THEOREM (Hölder's inequality). *Let X and Y be random variables and let $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$. If $X \in L^p$ and $Y \in L^q$ then*

$$\mathbb{E}(XY) \leq \mathbb{E}(|X|^p)^{1/p}\mathbb{E}(|Y|^q)^{1/q}$$

with equality if and only if $X = 0$ almost surely or $|Y| = a|X|^{p-1}$ almost surely for some constant $a \geq 0$. The case when $p = q = 2$ is called the Cauchy–Schwarz inequality.

DEFINITION. A random variable X is called *discrete* if X takes values in a countable set; i.e. there is a countable set S such that $X \in S$ almost surely. If X is discrete, the function $p_X : \mathbb{R} \to [0,1]$ defined by $p_X(t) = \mathbb{P}(X = t)$ is called the *mass function* of X.

The random variable X is *absolutely continuous* (with respect to Lebesgue measure) if and only if there exists a function $f_X : \mathbb{R} \to [0, \infty)$ such that

$$\mathbb{P}(X \leq t) = \int_{-\infty}^{t} f_X(x)dx$$

for all $t \in \mathbb{R}$, in which case the function f_X is called the *density function* of X.

If X is a random vector taking values in $\mathbb{R}^n$, then the density of X, if it exists, is the function $f_X : \mathbb{R}^n \to [0, \infty)$ such that

$$\mathbb{P}(X \in A) = \int_A f_X(x)dx$$

for all Borel subsets $A \subseteq \mathbb{R}^n$.

THEOREM. *Let the function $g : \mathbb{R} \to \mathbb{R}$ be such that $g(X)$ is integrable.*

If X is a discrete random variable with probability mass function p_X taking values in a countable set S then

$$\mathbb{E}(g(X)) = \sum_{t \in S} g(t)\ p_X(t).$$

If X is an absolutely continuous random variable with density function f_X then

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x)\ f_X(x)\ dx.$$

More generally, if X is a random vector valued in $\mathbb{R}^n$ with density f_X and $g : \mathbb{R}^n \to \mathbb{R}$ then

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}^n} g(x)\ f_X(x)\ dx.$$

4. Special distributions

Definition. Let X be a discrete random variable taking values in $\mathbb{Z}_+$ with mass function p_X.

The random variable X is called

- *Bernoulli* with parameter p if

$$p_X(0) = 1 - p \text{ and } p_X(1) = p.$$

  where $0 < p < 1$. Then $\mathbb{E}(X) = p$ and $\mathrm{Var}(X) = p(1-p)$.
- *binomial* with parameters n and p, written $X \sim \mathrm{bin}(n,p)$, if

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \text{ for all } k \in \{0, 1, \ldots, n\}$$

  where $n \in \mathbb{N}$ and $0 < p < 1$. Then $\mathbb{E}(X) = np$ and $\mathrm{Var}(X) = np(1-p)$.
- *Poisson* with parameter λ if

$$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda} \text{ for all } k = 0, 1, 2, \ldots$$

  where $\lambda > 0$. Then $\mathbb{E}(X) = \lambda$.
- *geometric* with parameter p if

$$p_X(k) = p(1-p)^{k-1} \text{ for all } k = 1, 2, 3, \ldots$$

  where $0 < p < 1$. Then $\mathbb{E}(X) = 1/p$.

Definition. Let X be a continuous random variable with density function f_X.
The random variable X is called

- *uniform* on the interval (a, b), written $X \sim \mathrm{unif}(a, b)$, if

$$f_X(t) = \frac{1}{b-a} \text{ for all } a < t < b$$

  for some $a < b$. Then $\mathbb{E}(X) = \frac{a+b}{2}$.
- *normal* or *Gaussian* with mean μ and variance σ^2, written $X \sim N(\mu, \sigma^2)$, if

$$f_X(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \text{ for all } t \in \mathbb{R}$$

  for some $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Then $\mathbb{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$.
- *exponential* with rate λ, if

$$f_X(t) = \lambda e^{-\lambda t} \text{ for all } t \geq 0$$

  for some $\lambda > 0$. Then $\mathbb{E}(X) = 1/\lambda$.

If X is a random vector valued in $\mathbb{R}^n$ with density

$$f_X(x) = (2\pi)^{-n/2} \det(V)^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)\cdot V^{-1}(x-\mu)\right)$$

for a positive definite $n \times n$ matrix V and vector $\mu \in \mathbb{R}^n$, then X is said to have the n-dimensional normal (or Gaussian) distribution with mean μ and variance V, written $X \sim N_n(\mu, V)$. Then $\mathbb{E}(X_i) = \mu_i$ and $\operatorname{Cov}(X_i, X_j) = V_{ij}$.

5. Conditional probability and expectation, independence

Definition. Let B be an event with $\mathbb{P}(B) > 0$. The *conditional probability* of an event A given B, written $\mathbb{P}(A|B)$, is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

The *conditional expectation* of X given B, written $\mathbb{E}(X|B)$, is

$$\mathbb{E}(X|B) = \frac{\mathbb{E}(X\mathbb{1}_B)}{\mathbb{P}(B)}.$$

Theorem (The law of total probability). *Let $B_1, B_2, \ldots$ be disjoint, non-null events such that $\bigcup_{i=1}^{\infty} B_i = \Omega$. Then*

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

for all events A.

Definition. Let $A_1, A_2, \ldots$ be events. If

$$\mathbb{P}\Big(\bigcap_{i\in I} A_i\Big) = \prod_{i\in I} \mathbb{P}(A_i)$$

for every finite subset $I \subset \mathbb{N}$ then the events are said to be *independent*.

Random variables $X_1, X_2, \ldots$ are called *independent* if the events $\{X_1 \le t_1\}, \{X_2 \le t_2\}, \ldots$ are independent. The phrase 'independent and identically distributed' is often abbreviated *i.i.d.*

Theorem. *If X and Y are independent and integrable, then*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

6. Probability inequalities

Theorem (Markov's inequality). *Let X be a positive random variable. Then*

$$\mathbb{P}(X \ge \epsilon) \le \frac{\mathbb{E}(X)}{\epsilon}$$

for all $\epsilon > 0$.

Corollary (Chebychev's inequality). *Let X be a random variable with $\mathbb{E}(X) = \mu$ and $\operatorname{Var}(X) = \sigma^2$. Then*

$$\mathbb{P}(|X-\mu| \ge \epsilon) \le \frac{\sigma^2}{\epsilon^2}$$

for all $\epsilon > 0$.

7. Characteristic functions

Definition. The *characteristic function* of a real-valued random variable X is the function $\phi_X : \mathbb{R} \to \mathbb{C}$ defined by

$$\phi_X(t) = \mathbb{E}(e^{itX})$$

for all $t \in \mathbb{R}$, where $i = \sqrt{-1}$. More generally, if X is a random vector valued in $\mathbb{R}^n$ then $\phi_X : \mathbb{R}^n \to \mathbb{C}$ defined by

$$\phi_X(t) = \mathbb{E}(e^{it\cdot X})$$

is the characteristic function of X.

Theorem (Uniqueness of characteristic functions). *Let X and Y be real-valued random variables with distribution functions F_X and F_Y. Let ϕ_X and ϕ_Y be the characteristic functions of X and Y. Then*

$$\phi_X(t) = \phi_Y(t) \text{ for all } t \in \mathbb{R}$$

if and only if

$$F_X(t) = F_Y(t) \text{ for all } t \in \mathbb{R}.$$

8. Fundamental probability results

Definition (Modes of convergence). Let $X_1, X_2, \ldots$ and X be random variables.

- $X_n \to X$ almost surely if $\mathbb{P}(X_n \to X) = 1$
- $X_n \to X$ in L^p, for $p \geq 1$, if $\mathbb{E}|X|^p < \infty$ and $\mathbb{E}|X_n - X|^p \to 0$
- $X_n \to X$ in probability if $\mathbb{P}(|X_n - X| > \epsilon) \to 0$ for all $\epsilon > 0$
- $X_n \to X$ in distribution if $F_{X_n}(t) \to F_X(t)$ for all points $t \in \mathbb{R}$ of continuity of F_X

Theorem. *The following implications hold:*

$$\left.\begin{array}{c} X_n \to X \text{ almost surely} \\ \text{or} \\ X_n \to X \text{ in } L^p, p \geq 1 \end{array}\right\} \Rightarrow X_n \to X \text{ in probability } \Rightarrow X_n \to X \text{ in distribution}$$

Furthermore, if $r \geq p \geq 1$ then $X_n \to X$ in $L_r \Rightarrow X_n \to X$ in L^p.

Definition. Let $A_1, A_2, \ldots$ be events. The term *eventually* is defined by

$$\{A_n \text{ eventually}\} = \bigcup_{N\in\mathbb{N}} \bigcap_{n\geq N} A_n$$

and *infinitely often* by

$$\{A_n \text{ infinitely often}\} = \bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} A_n.$$

[The phrase 'infinitely often' is often abbreviated *i.o.*]

Theorem (The first Borel–Cantelli lemma). *Let $A_1, A_2, \ldots$ be a sequence of events. If*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$$

then $\mathbb{P}(A_n$ infinitely often$) = 0$.

THEOREM (The second Borel-Cantelli lemma). *Let* $A_1, A_2, \ldots$ *be a sequence of* independent *events. If*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$$

then $\mathbb{P}(A_n \text{ infinitely often}) = 1$.

THEOREM (Monotone convergence theorem). *Let* $X_1, X_2, \ldots$ *be positive random variables with* $X_n \leq X_{n+1}$ *almost surely for all* $n \geq 1$*, and let* $X = \sup_{n \in \mathbb{N}} X_n$*. Then* $X_n \to X$ *almost surely and*

$$\mathbb{E}(X_n) \to \mathbb{E}(X).$$

THEOREM (Fatou's lemma). *Let* $X_1, X_2, \ldots$ *be positive random variables. Then*

$$\mathbb{E}(\liminf_{n \uparrow \infty} X_n) \leq \liminf_{n \uparrow \infty} \mathbb{E}(X_n).$$

THEOREM (Dominated convergence theorem). *Let* $X_1, X_2, \ldots$ *and* X *be random variables such that* $X_n \to X$ *almost surely. If* $\mathbb{E}(\sup_{n \geq 1} |X_n|) < \infty$ *then*

$$\mathbb{E}(X_n) \to \mathbb{E}(X).$$

THEOREM (A strong law of large numbers). *Let* $X_1, X_2, \ldots$ *be independent and identically distributed integrable random variables with common mean* $\mathbb{E}(X_i) = \mu$*. Then*

$$\frac{X_1 + \ldots + X_n}{n} \to \mu \text{ almost surely.}$$

THEOREM (Central limit theorem). *Let* $X_1, X_2, \ldots$ *be independent and identically distributed with* $\mathbb{E}(X_i) = \mu$ *and* $\mathrm{Var}(X_i) = \sigma^2$ *for each* $i = 1, 2, \ldots,$ *and let*

$$Z_n = \frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Then $Z_n \to Z$ *in distribution, where* $Z \sim N(0, 1)$.

$\mathbb{R}$	the set of real numbers
$\mathbb{R}_+$	the set of non-negative real numbers $[0, \infty)$
$\mathbb{N}$	the set of natural numbers $\{1, 2, \dots\}$
$\mathbb{C}$	the set of complex numbers
$\mathbb{Z}$	the set of integers $\{\dots, -2, -1, 0, 1, 2, \dots\}$
$\mathbb{Z}_+$	the set of non-negative integers $\{0, 1, 2, \dots\}$
A^c	the complement of a set A, $A^c = \{\omega \in \Omega, \omega \notin A\}$
F_X	the distribution function of a random variable X
p_X	the mass function of a discrete random variable X
f_X	the density function of an absolutely continuous random variable X
ϕ_X	the characteristic function of X
$\mathbb{E}(X)$	the expected value of the random variable X
$\text{Var}(X)$	the variance of X
$\text{Cov}(X, Y)$	the covariance of X and Y
$\mathbb{E}(X\|B)$	the conditional expectation of X given the event B
$a \wedge b$	$\min\{a, b\}$
$a \vee b$	$\max\{a, b\}$
a^+	$\max\{a, 0\}$
$\limsup_{n\uparrow\infty} x_n$	the limit superior of the sequence $x_1, x_2, \dots$
$\liminf_{n\uparrow\infty} x_n$	the limit inferior of the sequence $x_1, x_2, \dots$
$a \cdot b$	Euclidean inner (or dot) product in $\mathbb{R}^n$, $a \cdot b = \sum_{i=1}^n a_i b_i$
$\|a\|$	Euclidean norm in $\mathbb{R}^n$, $\|a\| = (a \cdot a)^{1/2}$
$X \sim \nu$	the random variable X is distributed as the probability measure ν
$\mathbb{1}_A$	the indicator function of the event A
$N(\mu, \sigma^2)$	the normal distribution with mean μ and variance σ^2
$N_n(\mu, V)$	the n-dimensional normal distribution with mean $\mu \in \mathbb{R}^n$ and variance $V \in \mathbb{R}^{n\times n}$
$\text{bin}(n, p)$	the binomial distribution with parameters n and p
$\text{unif}(a, b)$	the uniform distribution on the interval (a, b)
L^p	the set of random variables X with $\mathbb{E}\|X\|^p < \infty$

TABLE 1. Notation

Index