Advanced Financial Models

Michael R. Tehranchi

Contents

Financial mathematics as a subject is young (as compared to, say, number theory), but it is mature enough now that there has emerged some consensus on the notation, vocabulary and important results. These notes are an attempt to present many of the main ingredients of this theory, mainly concerning the pricing and hedging of derivative securities.

But before launching into the story, we will begin by acknowledging some of the real-world complications that will not be discussed at length hereafter.

1. Standing assumptions: complications we ignore

Unfortunately, actual financial markets are very complicated. Of course, in order to develop a systematic financial theory, it is prudent to concentrate on the essential features of these markets and ignore the less essential complications. Therefore, the theory that will be presented in these notes is concerned with the analysis of market models that have plenty of simplifying assumptions.

That is not to say that these complications are not important. Indeed, there is active ongoing research attempting to remove these simplifying assumptions from the canonical theory. Below is a list of these assumptions.

1.1. Dividends. The total stock of a publicly traded firm is divided into a fixed number N of shares. The owner of each share is then entitled to the fraction $1/N$ of the total profit of the firm.[1] A portion of the firm's profit is usually reinvested by management, for instance by building new factories, but the rest of the profit is paid out to the shareholders. In particular, the owner of each share of stock will receive periodically a dividend payment.

However, in this course,

$$\boxed{\text{we will usually assume that there are no dividend payments.}}$$

Actually, this assumption is not as terrible as it sounds. We will see see shortly how to adapt the theory developed for assets that pay no dividends to incorporate assets that have non-zero dividend payments.

1.2. Tick size. Financial markets usually have a smallest increment of price, the tick. (The tick refers back to the days when prices were quoted on ticker tape.) Indeed, the tick size can vary from market to market, and even for assets traded in the same market.

However, in this course,

$$\boxed{\text{we will assume that the tick size is zero.}}$$

This is a convenient assumption for those who prefer continuous mathematics to discrete. It is usually a harmless assumption, unless the prices of interest are very close to zero.

[1] Actually, things are even more complicated. For instance, stocks can be classified as either common or preferred, with implications on dividends, voting rights and claims on the firm's assets in case of bankruptcy. Also, the number N of shares outstanding is not necessarily fixed–firms may issue new shares to raise cash, or they might buy back shares raising the stock price to reward the shareholders.

1.3. Transactions costs. Financial transactions are processed by a string of middle men, each of whom charge a fee for their services. Usually the fee is nearly proportional to the size of the transaction.

However, in this course,

$$\boxed{\text{we will assume that there are no transactions costs.}}$$

This assumption is justified by by the fact that transactions costs are often very small relative to the size of typical transactions. But one must always remember that in some applications, it might not be wise to neglect these costs.

1.4. Short-selling constraints. In the real world, it is actually possible for someone to sell an asset that he does not own. The essential mechanism is to borrow a share of that asset from a broker, and then immediately to sell it to the market. This procedure is called short selling.

Brokers, however, place contraints on this behaviour. Indeed, they usually require collateral and charge a fee for their service. Furthemore, if the market price of the asset increases, or if the price of the collateral decreases, the broker may ask the short seller to put up even more collateral.

However, in this course,

$$\boxed{\text{we will assume that there are no short-selling constraints.}}$$

Indeed, the theory of discrete-time trading is cleaner without additional assumptions on the sizes of trades. But we will see that to overcome some technical problems in the theory of continuous-time trading, it will be natural to restrict trading to what are called admissible strategies.

1.5. Divisibility of assets. There is another real-world trading constraint of a rather technical nature. The smallest unit of stock is the share. A share cannot be further divided – it is generally impossible to buy half a share of a particular stock.

However, in this course,

$$\boxed{\text{we will assume that assets are infinitely divisible.}}$$

1.6. Bid-ask spread. Real-world trading is asymmetrical since the price to buy a share is usually higher than the price to sell it. The reason is that are two different ways to buy or sell an asset listed on an exchange: the limit order and the market order.

A limit buy order is an offer to buy a certain number of shares of the asset at a certain price. A limit sell order is defined similarly. The collection of unfilled limit orders is called the limit order book.

At any time, there is the highest price for which there is an order to buy the asset. This is called the bid price. The lowest price for which there is an order to sell is called the ask price. The bid/ask spread is the difference. Figure 1 illustrates the evolution of a hypothetical limit order book as various orders arrive and are filled.

A market order are instructions to execute a transaction at the best available price. In particular, if the market order is to buy, then the lowest limit sell order is filled first. Therefore, for small market buy orders, the per share price paid is the ask price. Similarly, if a market sell order arrives, then the highest limit buy order is filled first, and hence the per share price received is the bid price.

However, in this course,

$$\boxed{\text{we will assume that there are no bid-ask spreads.}}$$

This assumption is justified by the observation that in many markets, the spread is very small. However, in times of crisis, this assumption is not usually applicable, and hence the theory breaks down dramatically.

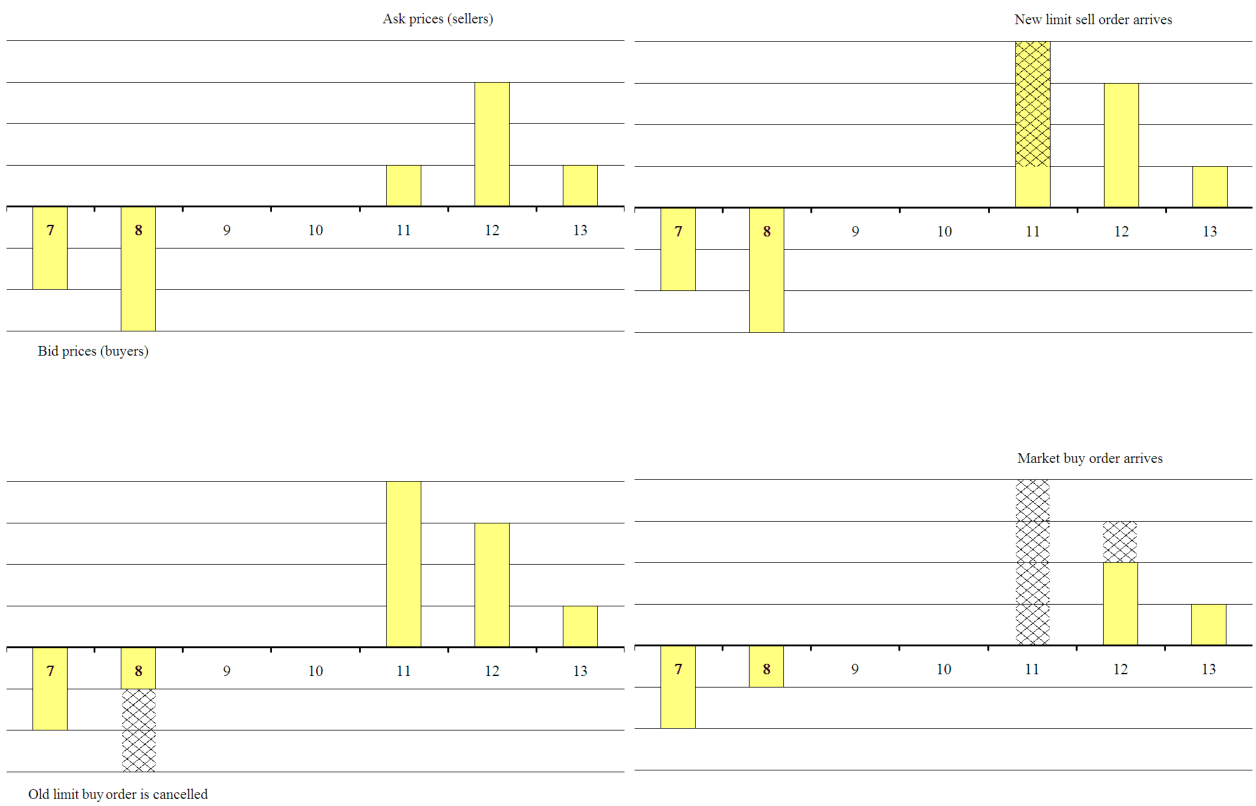


FIGURE 1. TOP LEFT. The bid price is £8 and the ask is £11. TOP RIGHT. A limit sell order for three shares at £11 arrives. BOTTOM LEFT. A limit buy order for two shares at £8 is cancelled. BOTTOM RIGHT. A market order to buy five shares arrives. Note that four shares are sold at £11 and one at £12. After the transaction, the ask price is £12.

1.7. Market depth. As described above, there are only a finite number of limit orders on the book at one time. If a large market buy order arrives, for instance, then the lowest limit sell order is filled first. But if the market order is bigger than the total shares available to buy at the ask price, then the limit orders at the next-to-lowest price are filled, and progresses up the book until the market order is finally filled. In this way, the ask price increases.

The market depth is the number of shares available to buy or sell at the ask or bid price respectively. Equivalently, the depth of a market is a measure of the size of a market order necessary to move quoted prices.

However, in this course,

$$\boxed{\text{we will assume that there is infinite market depth.}}$$

Equivalently, we will assume that investors are small relative to the limit order book, so they are price takers, not price makers. However, the most recent financial crisis shows that this assumption does not always approximate reality – just ask the traders at Lehman Brothers!

2. Further modelling complications

Microeconomic models usually involve the interaction of hypothetical agents who are endowed with preferences over some set of economic variables. The observed outcome of the system is then described by an equilibrium in which the competing preferences of the various agents are balanced through some economic mechanism, such as trade.

We will not deal much with such equilibrium models, but note here that equilibrium models make predictions about the structure of prices in a financial market (after we have made all of the simplifying assumptions listed above). In particular, in an equilibrium model, there cannot be an arbitrage. This will be explained in Chapter 1, but for the sake of this preface, we discuss some of the standard assumptions of equilibrium models and why they may fail in real life.

2.1. The expected utility hypothesis. In many standard microeconomic models, agents have preferences over random variables. For instance, suppose that the agent is young now but is planning for retirement. The amount of money that the agent will have in his pension fund when he retires can be modelled as a random variable. Of course, the particular random variable depends on the investment policy the agent chooses now.

The agent much choose his favourite investment strategy. Therefore, we must model his preferences over random variables. The *expected utility hypothesis* says that the agent prefers the random variable X to the random variable Y if and only if

$$\mathbb{E}[U(X)] > \mathbb{E}[U(Y)]$$

where the function $U : \mathbb{R} \to \mathbb{R}$, called the agent's *utility function* models the agents aversion risk.

The expected utility hypothesis seems to have a certain intuitive appeal, and practically, it does make the modelling problem more tractable. Furthermore, von Neumann and Morgenstern showed that the expected utility hypothesis is equivalent to a sensible seeming axiomisation of preferences.

2.2. The Allais paradox. The expected utility hypothesis can be tested, and it seems that real human beings do not always behave as though their preferences are consistent with it. Consider two games.

Game A. You must choose between a payment of either X or Y pounds, where

$$X = \begin{cases} 101 & \text{with prob. } 0.33 \\ 100 & \text{with prob. } 0.66 \\ 0 & \text{with prob. } 0.01 \end{cases} \quad \text{and } Y = 100 \text{ with prob.} 1$$

Game B. Again you must choose between a payment of either X or Y pounds, but now

$$X = \begin{cases} 100 & \text{with prob. } 0.34 \\ 0 & \text{with prob. } 0.66 \end{cases} \quad \text{and } Y = \begin{cases} 101 & \text{with prob. } 0.33 \\ 0 & \text{with prob. } 0.67 \end{cases}$$

Apparently, in real experiments, a significant number of people prefer Y in both games. For these people, their preferences are not compatible with the expected utility hypothesis. To see why not, suppose for the sake of finding a contradiction that the agent has a utility function U. Then

$$\text{Game A: } Y \text{ preferred } \Leftrightarrow 0.33\, U(101) + 0.66\, U(100) + 0.01\, U(0) < U(100)$$

and

$$\text{Game B: } Y \text{ preferred } \Leftrightarrow 0.34\, U(100) + 0.66\, U(0) < 0.33\, U(101) + 0.67\, U(0)$$

But the above inequalities cannot both hold true!

2.3. The Ellsberg paradox. Underlying the expected utility hypothesis is assumption that economic agents are perfect statisticians. In reality, of course, when faced with a random outcome, there is risk associated with the realisation of the randomness, but also uncertainty in the unknown distribution of the randomness. Here is an example.

Consider an urn with 30 balls, coloured red, yellow and black. You know that there are 10 red balls in the urn. However, you do not know the number of yellow balls or the number of black balls (but, of course, their sum to 20).

A single ball is drawn from the urn. Consider two games:

Game A.

$$X = \begin{cases} 100 & \text{if red} \\ 0 & \text{if yellow or black} \end{cases} \quad \text{and } Y = \begin{cases} 100 & \text{if yellow} \\ 0 & \text{if red or black} \end{cases}$$

Game B.

$$X = \begin{cases} 100 & \text{if red or black} \\ 0 & \text{if yellow} \end{cases} \quad \text{and } Y = \begin{cases} 100 & \text{if yellow or black} \\ 0 & \text{if red} \end{cases}$$

In experiments, people tend to prefer X in game A and Y in game B. This also contradicts the expected utility hypothesis. Suppose the agent estimates the probability of drawing yellow as p where $0 < p < 2/3$.

$$\text{Game A: } X \text{ preferred } \Leftrightarrow \frac{1}{3}\, U(100) + \frac{2}{3}\, U(0) > p\, U(100) + (1-p)\, U(0)$$

and

$$\text{Game B: } Y \text{ preferred } \Leftrightarrow (1-p)\, U(100) + p\, U(0) < \frac{2}{3}\, U(100) + \frac{1}{3}\, U(0)$$

a contradiction!

One way to avoid this paradox is to let the agent have preferences not only of risk but also uncertainty. For example, suppose that there is a probability models $\mathcal{P}$ consistent with the agent's beliefs. An agent who is very averse uncertainty would prefer X to Y if and only if

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}^{\mathbb{P}}[U(X)] > \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}^{\mathbb{P}}[U(Y)].$$

This point of view is pursued in an area called robust finance, but we will not discuss it in these notes.

3. Prerequisite knowledge

The emphasis of this course is on some of the mathematical aspects of financial market models. Very little is assumed of the reader's knowledge of the workings of financial markets. However, some mathematical background is needed.

Our starting point is the famous observation (sometimes attributed to Niels Bohr) that it is difficult to make predictions, especially about the future. Indeed, anyone with even a passing acquaintance with finance knows that most of us cannot predict with absolute certainty how the the price of an asset will fluctuate – otherwise we would be much richer!

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. Familiarity with measure theoretical probability is helpful, though a crashcourse on probability theory is given in an appendix.

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

CHAPTER 1

Discrete-time models

We consider a market with n assets. The identity of the assets is not important as long as the standing assumptions (zero dividends, zero tick size, zero transaction costs, no short-selling constraints, infinite divisibility, zero bid-ask spread, infinite market depth) are fulfilled. We usually think of the assets as being stocks and bonds, but they also can be more exotic things like pork belly futures.

We will use the notation P^i_t for the price of asset i at time t. In this section, the time index set is the non-negative integers, so the notation $t \geq 0$ should be interpreted as $t \in \{0, 1, 2, \dots\}$.

1. Measurability and conditional expectations

A modelling assumption that we will use throughout is that the collection of prices $P = (P^1_t, \dots, P^n_t)_{t\geq 0}$ is an n-dimensional stochastic process adapted to a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$. We now briefly describe what this means.

A stochastic process $(Z_t)_{t\geq 0}$ is just a collection of random variables (or vectors) indexed by the parameter t. In our case, the parameter is interpreted as time, either discrete or continuous. (Recall that when we speak of a random variable Z, we secretly have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ in the background such that the map $Z : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable. See the crashcourse if this sounds unfamiliar to you.)

We now formalise the concept of information being revealed as time marches forward. The correct notions are that of a filtration and adaptedness.

Motivation. You are probably already familiar with the notion of the measurability of a set. Measurability is a hugely important (though technical) idea in the theory of Lebesgue integration: for instance, Vitali showed that it is impossible to define the Lebesgue measure of *every* subset of $\mathbb{R}$.

But on top of its technical importance, measurability has is a way to model information. Give an probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\mathcal{G}$ be some set of information. For the moment, we will be vague about what $\mathcal{G}$ is, but it intuitively should have the property that an event $A \in \mathcal{F}$ is $\mathcal{G}$-measurable iff

$$\mathbb{P}(A|\mathcal{G}) \in \{0, 1\}.$$

Note we have not yet defined the notation $\mathbb{P}(A|\mathcal{G})$, but it should thought of as the probability of the event A given knowledge of the information set $\mathcal{G}$.

For instance, consider the experiment of tossing a fair coin two times. We can model this experiment on the sample space $\Omega = \{HH, HT, TH, TT\}$. The set of all events is the set

$$\mathcal{F} = \{\emptyset,\ \{HH\}, \dots, \{HH, HT\}, \dots, \{HH, HT, TH\}, \dots, \{HH, HT, TH, TT\}\}$$

of all $2^4 = 16$ subsets of Ω. The probability measure is just the one that assigns $\mathbb{P}(\{\omega\}) = 1/4$ equal probability to each elementary event.

Suppose $\mathcal{G}$ is information revealed by the first toss of the coin. The set event

$$A = \{ \text{ first toss is heads } \} = \{HH, HT\}$$

is $\mathcal{G}$-measurable, since for any sensible definition of the conditional probability we must have

$$\mathbb{P}(A|\mathcal{G}) = \begin{cases} 1 & \text{if the first toss is heads} \\ 0 & \text{if the first toss is tails.} \end{cases}$$

On the other hand, the event

$$B = \{ \text{ both tosses are heads } \} = \{HH\}$$

is not $\mathcal{G}$-measurable. Indeed, we must have

$$\mathbb{P}(B|\mathcal{G}) = \begin{cases} 1/2 & \text{if the first toss is heads} \\ 0 & \text{if the first toss is tails.} \end{cases}$$

Now returning to the general case, rather than modelling the set $\mathcal{G}$ of information as a new mathematical structure, we simply identify $\mathcal{G}$ with the collection of all $\mathcal{G}$-measurable events. Notice that, assuming that the conditional probability $\mathbb{P}(\cdot|\mathcal{G})$ somehow behaves like an unconditional probability, then $\mathcal{G}$ is a sigma-field. The lesson of all of this is that it makes sense to model information as a sub-sigma-field of the sigma-field of all events $\mathcal{F}$. It remains to properly define the conditional probability and then check that it has the correct properties.

We briefly recall some notions from probability.

Definition. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-sigma-field of events. A random variable $X : \Omega \to \mathbb{R}$ is *measurable* with respect to $\mathcal{G}$ (or briefly, $\mathcal{G}$-measurable) if and only if the event $\{X \leq x\}$ is an element of $\mathcal{G}$ for all $x \in \mathbb{R}$.

You know what that the conditional expectation of an integrable random variable X given a non-null event G means

$$\mathbb{E}(X|G) = \frac{\mathbb{E}(X \mathbb{1}_G)}{\mathbb{P}(G)}$$

The next theorem leads to a definition of conditional expectation given a sigma-field:

Theorem (Existence and uniqueness of conditional expectations). *Let X be an 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 $\mathcal{G} \subseteq \mathcal{F}$ 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 and $\bigcup_{n \in \mathbb{N}} G_n = \Omega$.

Let $\mathcal{G}$ be the smallest sigma-field containing $\{G_1, G_2, \ldots, ...\}$. That is, every element of $\mathcal{G}$ is of the form $\bigcup_{n \in I} G_n$ where $I \subseteq \mathbb{N}$. Then

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

where the right-hand side denotes conditional expectation given the event G_n.

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

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

Then

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

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.*
- *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.*
- *'slot property': 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*

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

Definition. The conditional probability of an event $A \in \mathcal{F}$ given a sub-sigma-field $\mathcal{G}$ is defined by

$$\mathbb{P}(A|\mathcal{G}) = \mathbb{E}(\mathbb{1}_A|\mathcal{G}).$$

We now come full circle to show that the motivation for defining measurability is compatible with the definitions we have chosen:

PROPOSITION. *If* $\mathbb{P}(A|\mathcal{G}) \in \{0,1\}$ *almost surely, then there exists a* $\mathcal{G}$*-measurable event* A' *such that* $\mathbb{P}(A \backslash A') = 0 = \mathbb{P}(A' \backslash A)$.

PROOF. Since the conditional probability takes values in $\{0,1\}$ there exists a $\mathcal{G}$-measurable event A' such that

$$\mathbb{P}(A|\mathcal{G}) = \mathbb{1}_{A'}.$$

Note that

$$\begin{aligned}
\mathbb{P}(A) =& \mathbb{E}(\mathbb{1}_A) \\
&= \mathbb{E}[\mathbb{E}(\mathbb{1}_A|\mathcal{G})] \qquad \text{tower} \\
&= \mathbb{E}[\mathbb{1}_{A'}] \\
&= \mathbb{P}(A')
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}(A \cap A') &= \mathbb{E}(\mathbb{1}_A \mathbb{1}_{A'}) \\
&= \mathbb{E}[\mathbb{E}(\mathbb{1}_A \mathbb{1}_{A'}|\mathcal{G})] \qquad \text{tower} \\
&= \mathbb{E}[\mathbb{1}_{A'} \mathbb{E}(\mathbb{1}_A|\mathcal{G})] \qquad \text{slot} \\
&= \mathbb{E}[\mathbb{1}_{A'}^2] \\
&= \mathbb{P}(A')
\end{aligned}$$

and hence

$$\mathbb{E}[(\mathbb{1}_A - \mathbb{1}_{A'})^2] = \mathbb{P}(A) + \mathbb{P}(A') - 2\mathbb{P}(A \cap A') = 0.$$

□

Continuing with the theme of measurability, we introduce a few more terms:

DEFINITION. A *filtration* $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a collection of sigma-fields such that $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for all $0 \leq s \leq t$.

DEFINITION. A process $X = (X_t)_{t \geq 0}$ is *adapted* to $\mathbb{F}$ iff the random variable X_t is $\mathcal{F}_t$-measurable for all $t \geq 0$.

When discussing an adapted stochastic process but a filtration is not explicitly mentioned, then we are implicitly working with the natural filtration of the process.

DEFINITION. Given a stochastic process $X = (X_t)_{t \geq 0}$, the *natural filtration* of X, (or the filtration *generated by* X) is the smallest filtration for which X is adapted. That is, it is the filtration $(\mathcal{F}_t)_{t \geq 0}$ where

$$\mathcal{F}_t = \sigma(X_s, 0 \leq s \leq t).$$

To gain some intuition about these definitions, consider this example.

EXAMPLE. Return to the experiment of tossing a fair coin two times. The flow of information is modelled by the following sigma-fields

- $\mathcal{F}_0 = \{\emptyset, \Omega\}$,
- $\mathcal{F}_1 = \{\emptyset, \{HH, HT\}, \{TH, TT\}, \Omega\}$,
- $\mathcal{F}_2 = \mathcal{F}$.

Now consider a stochastic process $(X_t)_{t\in\{0,1,2\}}$ that is adapted to the filtration $(\mathcal{F}_t)_{t\in\{0,1,2\}}$. Intuitively, the value of the random variable X_t is known once after t tosses of the coin.

For instance, X_0 must be a constant,

$$X_0(\omega) = a \text{ for all } \omega \in \Omega,$$

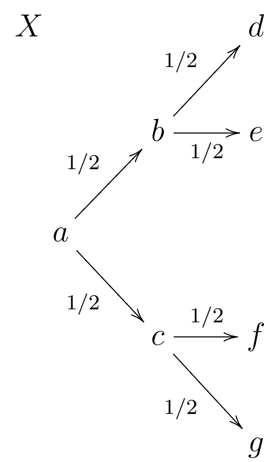
since there is no information before the experiment. On the other hand, the random variable X_1 must be of the form

$$X_1(\omega) = \begin{cases} b & \text{if } \omega \in \{HH, HT\} \\ c & \text{if } \omega \in \{TH, TT\} \end{cases}$$

since the only information known at time 1 is whether or not the first coin came up heads. Finally, X_2 can be any function on Ω, that is, of the form

$$X_2(\omega) = \begin{cases} d & \text{if } \omega = HH \\ e & \text{if } \omega = HT \\ f & \text{if } \omega = TH \\ g & \text{if } \omega = TT. \end{cases}$$

Alternatively, on this particular filtered probability space, the adapted process X can be visualised by the tree diagram:



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

For this course, it will be convenient to assume that there is no randomness at time 0. This can be made formal by assuming

$$\boxed{\text{the sigma-field } \mathcal{F}_0 \text{ is trivial.}}$$

This means that if A is an element $\mathcal{F}_0$ then either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. In particular, every $\mathcal{F}_0$-measurable random variable is almost surely constant. In the discrete-time theory, there nothing loss by further assuming $\mathcal{F}_0 = \{\emptyset, \Omega\}$. However, it turns out that this further assumption is technically inconvenient in the continuous-time theory.

Before continuing to the financial models, we list one final definition in this section.

DEFINITION. A *discrete-time* process $X = (X_t)_{t\ge1}$ is *previsible* (or *predictable*) iff the random variable X_t is $\mathcal{F}_{t-1}$-measurable for all $t \ge 1$.

Remark. Note that the time index set for a previsible process $(X_t)_{t\geq 1}$ is (usually) $\{1, 2, \ldots\}$, not $\{0, 1, \ldots\}$. In particular, X_0 is not necessarily defined.

Remark. In discrete time, a process X is previsible if and only if the process Y is adapted, where $X_t = Y_{t-1}$. That is to say, the notion of previsibility can be dispensed with by simply changing notation. However, in continuous time, there is a more subtle difference between the notions of previsibility and adaptedness. Therefore, for the sake of a unified treatment of the discrete and continuous time cases, we keep it in.

2. The set-up

Returning to our financial modelling, we assume that the market prices are given by a n-dimensional adapted process $P = (P_t)_{t\geq 0}$. For the moment, we assume that each asset may pay a dividend. Let δ_t^i be the dividend paid for each share of asset i held at time t. We assume that $(\delta_t)_{t\geq 1}$ is an adapted process. Note that we start time at $t = 1$ for this process - the reason will be come clear shortly. Also, recall from the preface that for most of the rest of the course we will assume that $\delta_t = 0$ almost surely for all $t \geq 1$. But to be clear, we do not make that assumption now.

To the market described by the adapted processes P and δ, we now introduce an investor. At time t, the investor comes into the market with an initial amount of money X_t. The investor receives some income I_t and consumes some amount C_t.

With the remaining money $X_t + I_t - C_t$, the investor buys a portfolio $H_{t+t} \in \mathbb{R}^n$ of the assets, where the real number H_{t+1}^i denotes the number of shares held in asset i. (If $H_{t+1}^i > 0$ then the position is said to be *long*, and if $H_{t+1}^i < 0$ then the position is said to be *short*.)

There are two important accounting relationships between the variables. First, the budget constraint is

$$X_t + I_t - C_t = H_{t+1} \cdot P_t,$$

where we are using the notation

$$a \cdot b = \sum_{i=1}^{n} a^i b^i$$

for the usual Euclidean inner (or dot) product in $\mathbb{R}^n$. Second, at time $t + 1$, the investor's portfolio is worth $X_{t+1} = H_{t+1} \cdot (P_{t+t} + \delta_{t+1})$.

Second, we have the accounting identity

$$X_{t+1} = H_{t+1} \cdot (P_{t+1} + \delta_{t+1})$$

which says that the amount of money the investor has in the market at time $t + 1$ is equal to the sum of liquidation value of the assets and the accrued dividends.

We will model the price process P, the dividend process δ and external income process I as exogenously given adapted processes. We consider the investor's initial wealth $X_0 = x$ as a given constraint, and the investor's portfolio process $(H_t)_{t\geq 1}$ as her control. In order to eliminate clairvoyant investors, we insist that the control H is previsibile.

Note that given x and H, the investor's wealth is

$$X_t^{x,H} = \begin{cases} x & \text{if } t = 0 \\ H_t \cdot (P_t + \delta_t) & \text{if } t \geq 1 \end{cases}$$

and the consumption is then

$$C_t^{x,H} = X_t^{x,H} + I_t - H_{t+1} \cdot P_t.$$

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. We consider one such optimal investment problem. The main motivation for studying this problem is to introduce the very important notion of a martingale deflator.

Let $T > 0$ be some non-random time horizon and prefers a consumption stream $c = (c_t)_{0 \le t \le T}$ to $c' = (c'_t)_{0 \le t \le T}$ iff and only if

$$\mathbb{E}[U(c)] > \mathbb{E}[U(c')]$$

where $U : \mathbb{R}^{1+T} \to \mathbb{R} \cup \{-\infty\}$ is a given utility function. We will assume that $c_t \mapsto U(c_0, \ldots, c_T)$ is strictly increasing for each t, modelling the assumption that the investor strictly prefers more to less. (Usually we also assume that U is strictly concave, so that the investor is risk-averse, strictly preferring to consume the non-random quantity $\mathbb{E}(c)$ to the random quantity c, for any non-constant random vector c.)

We suppose that investor's initial wealth is x given. We also suppose that he will live exactly to age T, and since he derives no utility from wealth in the afterlife, chooses to consume his remaining wealth at time T. Summing up, the investor faces the problem

$$\text{maximise } \mathbb{E}[U(c)] \text{ subject to } \begin{cases} c_t = C_t^{x,H} \text{ for } 0 \le t \le T \\ c_T = H_T \cdot (P_T + \delta_T) + I_T \end{cases}$$

With this problem in mind, we introduce an important definition:

With this problem in mind, we introduce an important definition:

Definition. An *arbitrage* is an n-dimensional previsible process H such that there exists a non-random time $T > 0$ with the properties, that the consumption stream

$$\begin{aligned} c_0 &= -H_1 \cdot P_0, \\ c_t &= H_t \cdot (P_t + \delta_t) - H_{t+1} \cdot P_t, \text{ for } 1 \le t \le T-1, \\ c_T &= H_T \cdot (P_T + \delta_T) \end{aligned}$$

satisfies

- $c_t \ge 0$ almost surely for all $0 \le t \le T$,
- $\mathbb{P}\,(c_t > 0 \text{ for some } 0 \le t \le T) > 0$.

Note that if H^{f} is a feasible investment strategy for the above investment problem and if H^{a} is an arbitrage, then $H^{\mathrm{f}} + H^{\mathrm{a}}$ is also feasible (since it can be funded with same initial wealth x). However, the new strategy has strictly higher expected utility

$$\mathbb{E}[U(c^{\mathrm{f}} + c^{\mathrm{a}})] > \mathbb{E}[U(c^{\mathrm{f}})].$$

Inductively, the strategy $H^{\mathrm{f}} + kH^{\mathrm{a}}$ is feasible for every $k \ge 0$. In particular, if there is an arbitrage *then there cannot be an optimal investment strategy* to the utility maximisation problem.

Aside. And why do we care about the existence of optimal investment strategies? To explain, we take a moment to ask where do prices come from? We consider a one-period model. Assume that that only happens at time 0, so we assign the terminal price vector

$P_1 = 0$. On the other hand, the vector of dividends δ_1 is unknown at time 0. It is possibly non-zero, but since its value is only revealed at time 1, we model it as a random vector. What determines the randomness? One could argue that all that matters is the *beliefs* of the market participants, not the underlying mechanism that causes the apparent randomness. So, we assume that there are J investors, and each investor j has a probability measure $\mathbb{P}_j$, where $j = 1, \ldots, J$ modelling the distribution of δ_1. We also assume that each agent j comes to the market with initial capital x_j. The market already has the n assets, with total supply of asset i given by S^i and $S = (S^1, \ldots, S^n)$. The agents trade with each other until each arrive at an optimal allocation H_j^* and collectively determine an initial price P_0^*. To formalise this, we have the following definition:

DEFINITION. Given initial wealths x_j, utility functions U_j and probability measures $\mathbb{P}_j$, for $j = 1, \ldots J$ which determine the distribution of the random vector P_1, let

$$H_j(p) = \arg\max\{\mathbb{E}_j[U(c_0, c_1)] : \quad c_0 = x_j - H \cdot p, \ c_1 = H \cdot \delta_1\}$$

be the optimal portfolio for agent i assuming the initial price is $P_0 = p$. An equilibrium price P_0^* is a solution to the equation

$$\sum_{j=1}^{J} H_j(P_0^*) = S,$$

where the notation $\mathbb{E}_j$ denotes expectation with respect to $\mathbb{P}_j$.

Note that the above condition says that for the equilibrium initial price P_0^*, the agents portfolios H_j^* solve their version of the optimal investment problem (*). A consequence of the previous proposition is the following motivating result:

PROPOSITION. *If the market is in equilibrium then no agent can believe there is an arbitrage.*

Before proceeding, we consider a simple consequence of the assumption of no arbitrage in a market model.

PROPOSITION. *Consider a market with $n = 1$ asset, and assume that $\delta_t \geq 0$ almost surely for all $t \geq 1$. If there exists a non-random $T > 0$ such that $P_T \geq 0$ almost surely, then $P_t \geq 0$ almost surely for all $0 \leq t \leq T$.*

PROOF. Let $\tau = \inf\{t \geq 0 : P_t < 0\}$ and let

$$H_t = \mathbb{1}_{\{\tau < t \leq T\}}$$

Note that H is previsible since for $1 \leq t \leq T$ we have

$$\{\tau < t\} = \cup_{s=0}^{t-1}\{P_s < 0\} \in \mathcal{F}_{t-1}.$$

Hence $c_t = -P_t > 0$ on $\{t = \tau < T\}$ and $c_T = P_T + \delta_T \geq 0$. Since there is no arbitrage, we must have $\mathbb{P}(\tau < T) = 0$, or equivalently, $\mathbb{P}(P_s \geq 0$ for all $0 \leq s \leq T-1) = 1$ as claimed. $\square$

3. The first fundamental theorem and martingales

We have tried to argue above that it is natural to insist that our market model is free of arbitrage strategies. But how can we check that a given price process P is arbitrage free? The answer is contained in a famous theorem:

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

We spend the next few pages unpacking this theorem. First, we need a definition to get started:

DEFINITION. A martingale deflator is an adapted (real-valued) process Y such that $Y_t > 0$ for all $t \geq 0$ almost surely, and such that the n-dimensional process

$$M_t = P_t Y_t + \sum_{s=1}^{t} \delta_s Y_s$$

is a martingale.

We now briefly review some bits of martingale theory.

*****

Now we come to one of the most important concepts in financial mathematics, the martingale. A martingale is simply an adapted stochastic process that is constant on average in the following sense:

DEFINITION. A *martingale* relative to a filtration $\mathbb{F}$ is an adapted stochastic process $M = (M_t)_{t \geq 0}$ with the following properties:

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

REMARK. The above definition of martingale is the same both discrete- and continuous-time processes. However, if the time index set is discrete $\mathbb{T} = \mathbb{Z}_+$, 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.

EXAMPLE. Let $\xi_1, \xi_2, \xi_3, \ldots$ be independent integrable random variables such that $\mathbb{E}(\xi_i) = 0$ for all i. The process $(S_t)_{t \geq 0}$ given by $S_0 = 0$ and

$$S_t = \xi_1 + \ldots + \xi_t$$

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

$$\mathbb{E}(|S_t|) \leq \mathbb{E}(|\xi_1|) + \ldots + \mathbb{E}(|\xi_t|)$$

by the triangular inequality and all the terms in this finite sum are finite by assumption. Also,

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

where the conditional expectation $\mathbb{E}(\xi_{t+1}|\mathcal{F}_t)$ is replaced by the unconditional expectation $\mathbb{E}(\xi_{t+1})$ by the assumption that ξ_{t+1} is independent of $\mathcal{F}_t = \sigma(S_1, \ldots, S_t) = \sigma(\xi_1, \ldots, \xi_t)$.

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 $M = (M_t)_{t \geq 0}$ is a martingale.

Integrability follows from the definition of conditional expectation. Now, for every $0 \leq s \leq t$ we have

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

by the tower property. Notice that this example also works in continuous time.

Sometimes we are given a process $(M_t)_{0 \leq t \leq T}$ where $T > 0$ is a fixed, non-random time horizon. To check that this process is a martingale, we need only check that

$$M_t = \mathbb{E}(M_T|\mathcal{F}_t) \text{ for all } 0 \leq t \leq T,$$

because this corresponds to the construction above with $X = M_T$.

This last example is theorem shows how to take one martingale and build another one.

Proposition. *Let M be a martingale and let K be a bounded predictable process. Then the process X defined by*

$$X_t = \sum_{s=1}^{t} K_s(M_s - M_{s-1})$$

is a martingale.

Proof. First, note X_t is $\mathcal{F}_t$-measurable by construction. Also, by assumption, we have $\mathbb{E}(|M_t|) < \infty$ for all t since M is a martingale and that there exist a constant $C > 0$ such that $|K_t| \leq C$ almost surely for all $t \geq 0$. Hence

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

Using the predictability of K and the slot property of conditional expectation, we have

$$\begin{aligned} \mathbb{E}(X_{t+1} - X_t|\mathcal{F}_t) &= \mathbb{E}(K_{t+1}(M_{t+1} - M_t)|\mathcal{F}_t) \\ &= K_{t+1}\mathbb{E}(M_{t+1} - M_t|\mathcal{F}_t) \\ &= 0 \end{aligned}$$

implying the martingale property $\mathbb{E}(X_{t+1}|\mathcal{F}_t) = X_t$, since X_t is $\mathcal{F}_t$-measurable. $\square$

Remark. The martingale X above is often called a *martingale transform* or a *discrete time stochastic integral*. As we will see, it is one of the key building blocks for the continuous time theory to come.

4. Local martingales

It turns out that to prove the 1FTAP, even in the easy direction, we need a little more technology.

With that introduction, we begin our study of local martingales. First we start with a definition.

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. Obviously, non-random times are stopping times. That is, if $\tau = t_0$ for some fixed $t_0 \geq 0$, then $\{\tau \leq t\} = \Omega$ if $t_0 \leq t$ and $\emptyset$ otherwise.

Example. Here is a typical example of a stopping time. Let $(Y_t)_{t \geq 0}$ be a discrete-time adapted process and let A be a Borel set (for instance, an interval). Then the random variable

$$\tau = \inf\{t \geq 0 : Y_t \in A\}$$

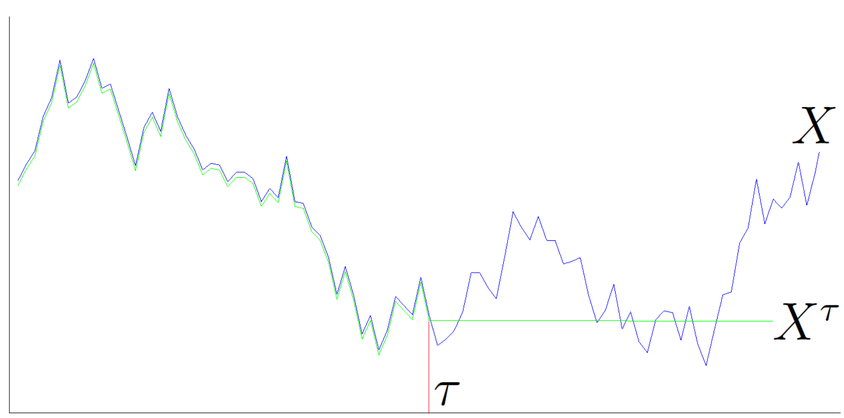
(with the usual convention that $\inf \emptyset = +\infty$) corresponding to the first time the process enters the set A is a stopping time. Indeed,

$$\{\tau \leq t\} = \bigcup_{s=0}^{t} \{Y_s \in A\}$$

is $\mathcal{F}_t$-measurable because each $\{Y_s \in A\}$ is $\mathcal{F}_s$-measurable by the adaptedness of Y, and $\mathcal{F}_s \subseteq \mathcal{F}_t$ by the definition of filtration.

Stopping times can be used to stop processes.

Definition. For an adapted process X (in discrete or continuous[1] time) and a stopping time τ, the process X^τ defined by $X^\tau_t = X_{t \wedge \tau}$ is said to be X stopped at τ.



Stopping times interact well martingales: stopped martingales are still martingales.

[1]If time is continuous, we also need the extra technical assumption that X is progressively measurable in order that the map $\omega \mapsto X_{\tau(\omega)}(\omega)$ is measurable. Fortunately, it is sufficient to assume that sample paths of X are continuous, which will be enough for this course.

Proposition. *Let X be a discrete-time martingale and let τ be a stopping time. Then X^τ is a martingale.*

Remark. A version of this theorem also holds for continuous-time martingales with continuous sample paths.

Proof. Note that

$$X_t^\tau = X_0 + \sum_{s=1}^{t} \mathbb{1}_{\{s \le \tau\}}(X_s - X_{s-1}).$$

Since the event $\{t \le \tau\} = \{\tau \le t-1\}^c$ is $\mathcal{F}_{t-1}$-measurable by the definition of stopping time, the process $K_t = \mathbb{1}_{\{t \le \tau\}}$ is predictable. Since X^τ is the martingale transform of the bounded predictable process K with respect to the martingale X, it is a martingale. □

The above result says that the martingale property is stable under stopping. We use this property as motivation for the following definition.

Definition. A *local martingale* is an adapted process $X = (X_t)_{t \ge 0}$, in either discrete or continuous time, such that there exists an increasing sequence of stopping times (τ_N) with $\tau_N \uparrow \infty$ such that the stopped process X^{τ_N} is a martingale for each N.

Remark. Note that martingales are local martingales. Indeed, given a martingale X and *any* sequence of stopping times $\tau_N \uparrow \infty$, the stopped process X^{τ_N} is a martingale.

Remark. Note that the local martingale property is also stable under stopping. Indeed, let X be a local martingale and τ a stopping time. Then by definition, there exists a sequence of stopping times $\sigma_N \uparrow \infty$ such that X^{σ_N} is a martingale. Hence $(X^{\sigma_N})^\tau = X^{\sigma_N \wedge \tau}$ is again a martingale since $\sigma_N \wedge \tau$ is a stopping time. But note that $X^{\sigma_N \wedge \tau} = (X^\tau)^{\sigma_N}$, implying that the sequence of stopping times $\sigma_N \uparrow \infty$ is such that $(X^\tau)^{\sigma_N}$ is a martingale. This means X^τ is a local martingale.

Theorem. *Suppose M is a discrete-time local martingale and K is a predictable process. Let*

$$X_t = \sum_{s=1}^{t} K_s(M_s - M_{s-1})$$

for $t \ge 1$. Then X is a local martingale.

Remark. This is the martingale transform as before, but now do not insist that K is bounded or that M is a true martingale. As a consequence, we cannot assert that X is a true martingale, merely a local martingale. The idea is that by localising, we can study the algebraic and measurability structure of the martingale transform without worrying about integrability issues.

Proof. Since M is a local martingale by assumption, there exists a sequence of stopping times $(\tau_n)_n$ with $\tau_n \uparrow \infty$ a.s. such that M^{τ_n} is a martingale.

Let $u_n = \inf\{t \ge 0 : |K_{t+1}| > n\}$ with the convention $\inf \emptyset = +\infty$. Note that since K is predictable we have

$$\{u_n \le t-1\}^c = \{u_n \ge t\} = \{|K_s| \le n \text{ for all } 0 \le s \le t\} \in \mathcal{F}_{t-1}$$

and hence that u_n is a stopping time with $u_n \uparrow \infty$.[2]

Finally, let $v_n = \tau_n \wedge u_n$. Note $v_n \uparrow \infty$ and v_n is a stopping time since $\{v_n \leq t\} = \{\tau_n \leq t\} \cup \{u_n \leq t\}$.

Now $M^{v_n} = (M^{\tau_n})^{u_n}$ is a stopped martingale, and hence a martingale. Also $(K_t \mathbb{1}_{\{t \leq v_n\}})_{t \geq 1}$ is a predictable process, bounded by n. Writing

$$X_t^{v_n} = \sum_{s=1}^{t} K_s \mathbb{1}_{\{s \leq v_n\}} (M_s^{v_n} - M_{s-1}^{v_n})$$

we see that the stopped process is the martingale transform of a bounded predictable process with respect to the martingale, and hence is a martingale. $\square$

The next theorem gives a sufficient condition that a local martingale is a true martingale.

Theorem. *Let X be a local martingale in either discrete or continuous time. Let Y_t be a process such that $|X_s| \leq Y_t$ almost surely for all $0 \leq s \leq t$. If $\mathbb{E}(Y_t) < \infty$ for all $t \geq 0$, then X is a true martingale.*

Proof. Let $(\tau_N)_N$ be a localising sequence of stopping times for X. Note that $X_{t \wedge \tau_N} \to X_t$ a.s. since $\tau_N \uparrow \infty$. Furthermore, by assumption $|X_{t \wedge \tau_N}| \leq Y_t$ which is integrable, so we may apply the conditional version of the dominated convergence theorem to conclude

$$\begin{aligned} \mathbb{E}(X_t | \mathcal{F}_s) &= \mathbb{E}(\lim_N X_{t \wedge \tau_N} | \mathcal{F}_s) \\ &= \lim_N \mathbb{E}(X_{t \wedge \tau_N} | \mathcal{F}_s) \\ &= \lim_N X_{s \wedge \tau_N} \\ &= X_s \end{aligned}$$

for $0 \leq s \leq t$, where we have used the fact that the stopped process $(X_{t \wedge \tau_N})_{t \geq 0}$ is a martingale. $\square$

The following corollary is useful:

Corollary. *Suppose X is a DISCRETE-TIME local martingale such that $\mathbb{E}(|X_t|) < \infty$ for all $t \geq 0$. Then X is a true martingale.*

Proof. Let $Y_t = |X_0| + \ldots + |X_t|$. The process Y is integrable by assumption and $|X_s| \leq Y_t$ for all $0 \leq s \leq t$. The conclusion follows from the previous theorem. $\square$

In the absence of integrability, the next best property is non-negativity. First we need some definitions.

Definition. A *supermartingale* relative to a filtration $(\mathcal{F}_t)_{t \geq 0}$ is an adapted stochastic process $(U_t)_{t \geq 0}$ with the following properties:

[2] Note that

$$\begin{aligned} \{\sup_n u_n \geq t\} &= \cup_n \{u_n \geq t\} \\ &= \cup_n \cap_{s=1}^{t} \{|K_s| \leq n\} \\ &= \{\max_{1 \leq s \leq t} |K_t| < \infty\} \end{aligned}$$

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

A *submartingale* is an adapted process $(V_t)_{t\geq 0}$ with the following properties:

- $\mathbb{E}(|V_t|) < \infty$ for all $t \geq 0$
- $\mathbb{E}(V_t|\mathcal{F}_s) \geq V_s$ for all $0 \leq s \leq t$.

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.

As in the case of the definition of martingale, to show that an adapted, integrable process U is a supermartingale in discrete time, it is enough to show that $\mathbb{E}(U_{t+1}|\mathcal{F}_t) \leq U_t$ for all $t \geq 0$.

THEOREM. *Suppose X is a local martingale in either continuous or discrete time. If $X_t \geq 0$ for all $t \geq 0$, then X is a supermartingale.*

PROOF. In the general case, let $(\tau_N)_N$ be the localising sequence for X. First we show that X_t is integrable for each $t \geq 0$. Fatou's lemma yields

$$\begin{aligned}\mathbb{E}(|X_t|) &= \mathbb{E}(X_t)\\ &= \mathbb{E}(\lim_N X_{t\wedge\tau_N})\\ &\leq \liminf_N \mathbb{E}(X_{t\wedge\tau_N})\\ &= X_0 < \infty.\end{aligned}$$

Now that we have established integrability, we can discuss conditional expectations. The conditional version of Fatou's lemma yields

$$\begin{aligned}\mathbb{E}(X_t|\mathcal{F}_s) &= \mathbb{E}(\lim_N X_{t\wedge\tau_N}|\mathcal{F}_s)\\ &\leq \liminf_N \mathbb{E}(X_{t\wedge\tau_N}|\mathcal{F}_s)\\ &= \liminf_N X_{s\wedge\tau_N}\\ &= X_s\end{aligned}$$

for $0 \leq s \leq t$, as claimed. □

As before, discrete time local martingales are particularly nice:

COROLLARY. *If X is a DISCRETE-TIME local martingale such that $X_t \geq 0$ a.s. for all $t \geq 0$, then X is a martingale.*

PROOF. By the above theorem, we have that $\mathbb{E}(|X_t|) = \mathbb{E}(X_t) \leq X_0 < \infty$. Since X is integrable, the previous corollary implies X is a martingale. □

THEOREM. *Suppose that*

$$X_t = X_0 + \sum_{s=1}^{t} K_s(M_s - M_{s-1})$$

where K is predictable, X is a martingale and X_0 is a constant. If $X_T \geq 0$ a.s. for some non-random $T > 0$, then $(X_t)_{0\leq t\leq T}$ is a true martingale.

PROOF. Just as before, let $\tau_N = \inf\{t \geq 0 : |K_{t+1}| > N\}$. Note $X_s \mathbb{1}_{\{t \leq \tau_N\}}$ is integrable for all $0 \leq s \leq t$, since M is integrable by definition of martingale, and K_s is bounded on $\{t \leq \tau_N\}$. Hence we have

$$\begin{aligned}
0 &\leq \mathbb{E}[X_T \mathbb{1}_{\{T \leq \tau_N\}} | \mathcal{F}_{T-1}] \\
&= \mathbb{E}[X_{T-1} \mathbb{1}_{\{T \leq \tau_N\}} + K_T \mathbb{1}_{\{T \leq \tau_N\}} (M_T - M_{T-1}) | \mathcal{F}_{T-1}] \\
&= X_{T-1} \mathbb{1}_{\{T \leq \tau_N\}} + K_T \mathbb{1}_{\{T \leq \tau_N\}} \mathbb{E}[M_T - M_{T-1} | \mathcal{F}_{T-1}] \\
&= X_{T-1} \mathbb{1}_{\{T \leq \tau_N\}}.
\end{aligned}$$

Taking $N \to \infty$ shows $X_{T-1} \geq 0$ a.s., induction shows that $X_t \geq 0$ for all $0 \leq t \leq T$. Therefore $(X_t)_{0 \leq t \leq T}$ is a non-negative local martingale in discrete time and hence a true martingale. □

5. Proof of the 1FTAP, easier direction

Recall our framework. There exist n-dimensional price $(P_t)_{t \geq 0}$ and dividend $(\delta_t)_{t \geq 1}$ processes, adapted to a given filtration $(\mathcal{F}_t)_{t \geq 0}$. A martingale deflator is a positive (real-valued) adapted process $(Y_t)_{t \geq 0}$ such that YP and $Y\delta$ are integrable and

$$\mathbb{E}[Y_{t+1}(P_{t+1} + \delta_{t+1}) | \mathcal{F}_t] = Y_t P_t \text{ for all } t \geq 0.$$

Our aim is to show that if there exists a martingale deflator, then there is no arbitrage. We first prove a useful lemma.

LEMMA. *Given a real constant x and a n-dimensional previsible process $(H_t)_{t \geq 1}$, let*

$$\begin{aligned}
X_0 &= x \\
X_t &= H_t \cdot (P_t + \delta_t) \textit{ for } t \geq 1,
\end{aligned}$$

and

$$c_t = X_t - H_{t+1} \cdot P_t \quad \textit{for } t \geq 0.$$

Suppose there exists a martingale deflator Y. Set

$$Z_t = X_t Y_t + \sum_{s=0}^{t-1} c_s Y_s.$$

Then M is a local martingale. Furthermore, if $c_t \geq 0$ a.s for $0 \leq t \leq T-1$ and $X_T \geq 0$ a.s for some non-random $T > 0$, then $(M_t)_{0 \leq t \leq T}$ is a martingale.

PROOF. Note that

$$\begin{aligned}
Z_t - Z_{t-1} &= X_t Y_t + Y_{t-1}(c_{t-1} - X_{t-1}) \\
&= H_t \cdot [Y_t(P_t + \delta_t) - Y_{t-1} P_{t-1}] \\
&= H_t \cdot (M_t - M_{t-1})
\end{aligned}$$

where

$$M_t = P_t Y_t + \sum_{s=1}^{t} \delta_s Y_s.$$

Note that Z is martingale transform of the previsible process H with respect to the martingale M, and hence Z is a local martingale.

Furthermore, if $Z_T \geq 0$ a.s., from the last theorem of the previous section, we have $(Z_t)_{0 \leq t \leq T}$ is a true martingale $\square$

Proof of the 1FTAP, easier direction. Suppose that there is a martingale deflator Y. Let H be an n-dimensional previsible process and let $c_0 = -H_1 \cdot P_0$ and

$$c_t = H_t \cdot (P_t + \delta_t) - H_{t+1} \cdot P_t \text{ for } t \geq 1.$$

Suppose there is some non-random $T > 0$ such that $c_T = H_T \cdot (P_T + \delta_T)$ and $c_t \geq 0$ almost surely for all $0 \leq t \leq T$. To show that H is not an arbitrage, must show that $c_t = 0$ almost surely for all $0 \leq t \leq T$.

To this end, let

$$Z_T = \sum_{s=0}^{T} c_s Y_s.$$

Since $Y_s > 0$ and $c_s \geq 0$ for all $0 \leq s \leq T$, we need only show that $Z_T = 0$ almost surely. By the pigeon-hole principle, it is sufficient to show

$$\mathbb{E}(Z_T) = 0.$$

To finish the proof, let

$$Z_t = X_t Y_t + \sum_{s=0}^{t-1} c_s Y_s \text{ for } 0 \leq t \leq T$$

where we have set $X_0 = 0$ and $X_t = H_t \cdot (P_t + \delta_t)$ for $1 \leq t \leq T$. By the previous lemma, the process $(Z_t)_{0 \leq t \leq T}$ is a martingale. Since $Z_0 = 0$, we are done. $\square$

6. Proof of harder direction of the 1FTAP

In this section we will present elements of the proof of the first fundamental theorem of asset pricing. We will give a complete proof of the one-period case, and sketch the main steps to prove the full multi-period case. But first, we take a moment to reflect on the economic motivation.

6.1. Motivation: Langrangian duality. Consider the investor's utility maximisation problem to find an investment strategy H to

$$\text{maximise } \mathbb{E}[U(c)] \text{ subject to } \begin{cases} c_0 = x - H_1 \cdot P_0, \\ c_t = H_t \cdot (P_t + \delta_t) - H_{t+1} \cdot P_t \text{ for } 1 \leq t \leq T-1 \\ c_T = H_T \cdot (P_T + \delta_T) \end{cases}$$

where the utility function u is increasing in each argument. We have discussed previously that the existence of a maximiser implies that there does not exist an arbitrage. We now explain why the existence of a maximiser also points to the existence of a martingale deflator.

As usual in a constrained optimisation problem, we apply the Lagrangian method. Recall that this involves replacing our given objective function with the so-called Lagrangian which encodes the constraints. In this case the Lagrangian is

$$\begin{aligned}L(c,H,Y) =&\mathbb{E}[u(c_0,\dots,c_T)] + Y_0(x - c_0 - H_1\cdot P_0)\\&+\mathbb{E}\left[\sum_{t=1}^{T-1} Y_t[H_t\cdot(P_t+\delta_t) - H_{t+1}\cdot P_t - c_t] + Y_T[H_T\cdot(P_T+\delta_T) - c_T]\right]\end{aligned}$$

where the real-valued adapted process $(Y_t)_{0\le t\le T}$ is the family of Lagrange multipliers.

To identify the dual feasibility condition, we seek to find conditions on the Lagrange multiplier process Y implied by the existence of a maximiser of the Lagrangian $L(c,H,Y)$ over adapted c and previsible H. To this end, we rewrite the Lagrangian as

$$\begin{aligned}L(c,H,Y) =&\mathbb{E}\left[u(c_0,\dots,c_T) - \sum_{t=0}^{T} Y_t c_t\right] + xY_0\\&+\sum_{t=1}^{T}\mathbb{E}\{H_t\cdot[(P_t+\delta_t)Y_t - P_{t-1}Y_{t-1}]\}.\end{aligned}$$

By formally differentiating with respect to the t-th consumption variable, we find the maximised consumption c^* satisfies

$$\mathbb{E}\left[\frac{\partial u}{\partial c_t}(c^*)|\mathcal{F}_t\right] = Y_t.$$

Differentiating with respect to H_t yields

$$\mathbb{E}[(P_t+\delta_t)Y_t|\mathcal{F}_{t-1}] = P_{t-1}Y_{t-1}$$

These two conditions suggest that Y is an adapted, positive process such that the process

$$M_t = Y_tP_t + \sum_{s=1}^{t} Y_s\delta_s$$

is a martingale – i.e. Y is a martingale deflator.

6.2. Proof when $T = 1$. We now proceed to turn those fuzzy heuristics into a proper proof. We consider the one-period case. So our market data are the prices P_0, P_1 and the dividend δ_1. Since our calculations only involve $P_1+\delta_1$, but not the two terms separately, there is no loss just assuming $\delta_1 = 0$.

We suppose that the market $(P_t)_{t\in\{0,1\}}$ has no arbitrage, so that for any vector $H\in\mathbb{R}^n$ that has the property that $H\cdot P_0 \le 0 \le H\cdot P_1$ almost surely, it must be the case that $H\cdot P_0 = 0 = H\cdot P_1$ almost surely. We will show that, given any positive random variable Z, there exists a martingale deflator Y_0, Y_1 such that the product Y_1Z is bounded by a constant. This extra boundedness assumption is much stronger than what we need, but it comes for free from the proof and we will find it useful later in the course.

Let

$$\zeta = \frac{e^{-\|P_1\|^2/2}}{1+Z}.$$

Define a function $F:\mathbb{R}^n\to\mathbb{R}$ by

$$F(H) = e^{H\cdot P_0} + \mathbb{E}[e^{-H\cdot P_1}\zeta].$$

The positive random variable ζ is introduced to ensure integrability. Indeed note that the integrand $e^{-H\cdot P_1}\zeta \le e^{\|H\|^2/2}$ is bounded for each choice of H. In particular, the function F is finite everywhere and (by the dominated convergence theorem) smooth.

We will show that no investment-consumption arbitrage implies that the function F has a minimiser H^*. By the first order condition for a minimum, we have

$$0 = \nabla F(H^*) = e^{H^*\cdot P_0} P_0 - \mathbb{E}[e^{-H^*\cdot P_1}\zeta P_t]$$

and hence we may take

$$Y_0 = e^{H^*\cdot P_0} \text{ and } Y_1 = e^{-H^*\cdot P_1}\zeta.$$

Note that $Y_1 Z < C$ for some constant $C > 0$ (which depends on H^* in general).

So let $(H_k)_k$ be a sequence such that $F(H_k) \to \inf_H F(H)$. If $(H_k)_k$ is bounded, we can pass to a convergent subsequence, by the Bolzano–Weierstrass theorem, such that $H_k \to H^*$. By the smoothness of F we have

$$\inf_H F(H) = \lim_k F(H_k) = F(\lim_k H_k) = F(H^*)$$

so H^* is our desired minimiser.

It remains to show that no arbitrage implies that there exists a bounded minimising sequence $(H_k)_k$. We show that if every minimising sequence $(H_k)_k$ is unbounded, then there would be a contradiction.

Now we arrive at a little technicality. Let

$$\mathcal{U} = \{u \in \mathbb{R}^n : u \cdot P_0 = 0 = u \cdot P_1 \text{ a.s.}\} \subseteq \mathbb{R}^n$$

and let

$$\mathcal{V} = \mathcal{U}^\perp.$$

Notice that if $u \in \mathcal{U}$ and $v \in \mathcal{V}$ then $F(u + v) = F(v)$. Hence, by projecting a given minimising sequence onto the subspace $\mathcal{V}$, it is sufficient to consider minimising sequence $(H_k)_k$ taking values in $\mathcal{V}$.

So suppose, for the sake of finding a contradiction, that all minimising sequence $(H_k)_k$ taking values in $\mathcal{V}$ are unbounded. We can pass to a subsequence such that $\|H_k\| \uparrow \infty$. Now let

$$\hat{H}_k = \frac{H_k}{\|H_k\|}.$$

Note that $\|\hat{H}_k\| = 1$ and that $\hat{H}_k \in \mathcal{V}$. Since $(\hat{H}_k)_k$ is bounded, we can again pass to a convergent subsequence such that $\hat{H}_k \to \hat{H}$. Notice once more that $\|\hat{H}\| = 1$ and that $\hat{H} \in \mathcal{V}$.

We know that the sequence $F(H_k)$ is bounded (since it is convergent) but we also have

$$F(H_k) = (e^{\hat{H}_k\cdot P_0})^{\|H_k\|} + \mathbb{E}[(e^{-\hat{H}_k\cdot P_1})^{\|H_k\|}\zeta]$$

so we must conclude that $\hat{H} \cdot P_0 \le 0 \le \hat{H} \cdot P_1$ a.s. (since otherwise the right-hand side would blow up). By no-arbitrage, we conclude that the candidate arbitrage $\hat{H}$ is not actually an arbitrage, so $\hat{H} \cdot P_0 = 0 = \hat{H} \cdot P_1$ a.s.

We have shown that $\hat{H}$ is in $\mathcal{U}$. But since $\hat{H}$ is also in $\mathcal{V} = \mathcal{U}^\perp$, so we have $\hat{H} = 0$. And, finally, this contradicts $\|\hat{H}\| = 1$. $\square$

6.3. Elements of the proof of the harder direction of the multi-period 1FTAP. We have already seen the one period case. The full multi-period proof is a little more difficult because of some technicalities involving measurability.

We begin with two propositions that show that two of the existential-type results we needed in one-period proof have measurable versions.

PROPOSITION. *Let $f : \mathbb{R}^n \times \Omega \to \mathbb{R}$ be such that $f(x, \cdot)$ is measurable for all x, and that $f(\cdot, \omega)$ is continuous and has a unique minimiser $X^*(\omega)$ for each ω. Then X^* is measurable.*

Let us pause to think about what it means for the unique minimiser $x^* \in \mathbb{R}^n$ of a continuous function $g : \mathbb{R}^n \to \mathbb{R}$ to be in a closed rectangle of the form $A = [a_1, b_1] \times \cdots \times [a_n, b_n]$.

First note that that for all $q \neq x^*$, we have $g(x^*) < g(q)$. Let $Q \subset \mathbb{R}^n$ be countable and dense - for instance, let Q be the set of points with rational coordinates. By the continuity of g and the density of Q, we have that for every $q \neq x^*$, there exists a $p \in Q$ such that $g(p) < g(q)$.

Now if $x^* \in A$, then for any $q \in A^c \cap Q$, there exists a $p \in A \cap Q$ such that $g(p) < g(q)$, since $q \neq x^*$ and $A \cap Q$ is dense in A.

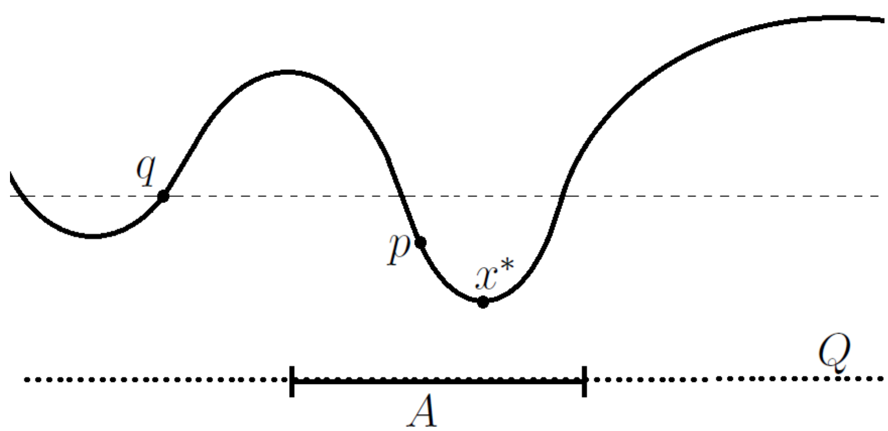
Conversely, suppose that for any $q \in A^c \cap Q$, there exists a $p \in A \cap Q$ such that $g(p) < g(q)$. This means

$$\inf_{x \in A \cap Q} g(x) \leq \inf_{x \in A^c \cap Q} g(x).$$

By the continuity of g the above inequality implies

$$\inf_{x \in A} g(x) = \inf_{x \in A \cup A^c} g(x) = g(x^*)$$

and in particular, we have $x^* \in A$ since A is closed.



PROOF. For any closed rectangle $A \subset \mathbb{R}^n$ we have

$$\{\omega : X^*(\omega) \in A\} = \bigcap_{q \in A^c \cap Q} \bigcup_{p \in A \cap Q} \{\omega : f(p, \omega) < f(q, \omega)\}$$

where Q is a countable dense subset of $\mathbb{R}^n$. Since the Borel sigma-field is generated by such rectangles, this implies the measurability of X^*. $\square$

We also need a useful measurable version of the Bolzano–Weierstrass theorem.

Proposition. *Let $(\xi_i)_{i\geq 1}$ be a sequence of measurable functions $\xi_i : \Omega \to \mathbb{R}^n$ such that $\sup_i \|\xi_i(\omega)\| < \infty$ for all $\omega \in \Omega$. Then there exists an increasing sequence of integer-valued measurable functions I_j and an $\mathbb{R}^n$-valued measurable function ξ^* such that*

$$\xi_{I_j(\omega)}(\omega) \to \xi^*(\omega) \text{ as } j \to \infty$$

for all $\omega \in \Omega$.

Proof. First we consider the $n = 1$ case. Let $\xi^*(\omega) = \limsup_i \xi_i(\omega)$. Note ξ^* is finite-valued and measurable, and that for every $j > 0$ there exists an infinite number of i's such that $\xi_i(\omega) \geq \xi^*(\omega) - 1/j$. Now let

$$I_j = \inf\{i \geq j : \xi_i \geq \xi^* - 1/j\}.$$

Since we have the representation of the event

$$\{I_j \leq h\} = \cup_{i=j}^{h}\{\xi_i \geq \xi^* - 1/j\}$$

for each $h \geq j$, the function I_j is measurable and $\xi_{I_j} \to \xi^*$ as desired.

Now we prove the claim for any dimension $n \geq 1$ by induction. Suppose that the claim is true for dimension $n = N$. Let $(\xi_i)_i$ be a sequence of measurable function valued in $\mathbb{R}^{N+1}$ such that $\sup_i \|\xi_i(\omega)\| < \infty$. Writing $\xi_i = (\zeta_i, \eta_i)$ where ζ_i takes values in $\mathbb{R}^N$ and η_i takes values in $\mathbb{R}$, we have by assumption the existence of a measurable sequence I_j and a measurable ζ^* such that

$$\zeta_{I_j} \to \zeta^*.$$

Notice that $(\eta_{I_j(\omega)}(\omega))_j$ is bounded for each ω, and hence by the $n = 1$ case, there exists an increasing measurable sequence J_k and a measurable η^* such that $\eta_{I_{J_k}} \to \eta^*$. In particular,

$$\xi_{I_{J_k}} \to (\zeta^*, \eta^*) = \xi^*$$

as desired. □

7. Numéraires and equivalent martingale measures

For the previous sections, we have worked in considerable generality. We have allowed our assets to pay a dividend, and we have even allowed the prices and dividends to be negative. In this section, we will see the effect of adding more assumptions to our model.

The first assumption that we will make from here on is that, unless otherwise indicated, there are no dividends:

$$\boxed{\textbf{Assumption:}\ \ \delta_t = 0 \text{ almost surely for all } t \geq 1}$$

Removing dividends from the conversation simplifies things somewhat. For instance, in this case, a positive adapted process Y is a martingale deflator if and only if the process YP is a martingale. It is sometimes useful (for example, for the example sheet) to introduce some vocabulary:

Definition. In a market with no dividends, a *pricing kernel* (or *stochastic discount factor* or *state price density*) between times s and t, where $0 \leq s < t$, is a positive $\mathcal{F}_t$-measurable random variable $\rho_{s,t}$ such that

$$P_s = \mathbb{E}(\rho_{s,t} P_t | \mathcal{F}_s).$$

Let Y be a martingale deflator, so that $\mathbb{E}(Y_t P_t|\mathcal{F}_s) = Y_s P_s$ for all $0 \leq s < t$, and let $\rho_{s,t} = Y_t/Y_s$. If $\rho_{s,t} P_t$ is integrable, then $\rho_{s,t}$ is a pricing kernel between times s and t. Conversely, suppose $\rho_{s,s+1}$ is a pricing kernel between times s and $s+1$ for $s \geq 0$, and let $Y_t = \rho_{0,1} \cdots \rho_{t-1,t}$. Then assuming YP is integrable, then Y is a martingale deflator.

Hopefully, after several sections where the theory is worked out in full generality, it is not too difficult to reincorporate dividends into the remaining bits of the course if desired.

The point of this section is now to consider the effect of adding the further assumption that there exists a portfolio with a strictly positive price.

Definition. A numéraire portfolio is n-dimensional previsible process $\eta = (\eta_t)_{t\geq 0}$ such that $\eta_t \cdot P_t > 0$ almost surely for each $t \geq 0$, and satisfying the pure-investment self-financing condition

$$(\eta_t - \eta_{t+1}) \cdot P_t = 0.$$

A numéraire asset is an asset with a strictly positive price.

Remark. If asset i is a numéraire asset, then the constant portfolio η defined for all $t \geq 0$ by

$$\eta_t^j = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

is a numéraire portfolio.

In the context of a market with a numéraire, we consider a different type of arbitrage:

Definition. A terminal consumption arbitrage is a n-dimensional previsible process H such that there exists a non-random $T > 0$ such that almost surely we have

$$\begin{aligned} c_0 &= -H_1 \cdot P_0 = 0 \\ c_t &= (H_t - H_{t+1}) \cdot P_t = 0 \text{ for } 1 \leq t \leq T-1 \\ c_T &= H_T \cdot P_T \geq 0 \end{aligned}$$

and

$$\mathbb{P}(c_T > 0) > 0.$$

Note that our first definition of arbitrage allowed for consumption at intermediate times.

Clearly, a terminal consumption arbitrage is an arbitrage according to our earlier definition. We now ask when does the existence of an arbitrage (possibly with intermediate consumption) imply the existence of a terminal consumption arbitrage.

We come to the answer to our question:

Proposition. *Consider a market with a numéraire η and let $N_t = \eta_t \cdot P_t$ for $t \geq 0$. Let H be an investment-consumption strategy with consumption stream*

$$\begin{aligned} c_0 &= x - H_1 \cdot P_0 \\ c_t &= (H_t - H_{t+1}) \cdot P_t, \end{aligned}$$

where x is the initial wealth. Let

$$K_t = H_t + \eta_t \sum_{s=0}^{t-1} \frac{c_s}{N_s}$$

for $t \geq 1$. Then K is a pure-investment strategy from the same initial wealth x. (That is, K be the strategy that consists of holding at time t the portfolio H_t but of instead of consuming the amount c_t, instead invest this money into the numéraire portfolio.)

In particular, H is an arbitrage if and only if K is a terminal consumption arbitrage.

Proof. Note that

$$\begin{aligned}(K_t - K_{t+1}) \cdot P_t =& (H_t - H_{t+1}) \cdot P_t - \eta_{t+1} \cdot P_t \frac{c_t}{N_t} \\ &+ (\eta_t - \eta_{t+1}) \cdot P_t \sum_{s=1}^{t-1} \frac{c_s}{N_s} \\ =& 0\end{aligned}$$

so K is a pure investment strategy by the assumption that η is pure-investment. Suppose $c_T = H_T \cdot P_T$ for some non-random time T. Then

$$K_T \cdot P_T = N_T \sum_{s=0}^{T} \frac{c_s}{N_s} \geq 0.$$

The left-hand side is positive if and only if c_t is positive for some $0 \leq t \leq T$. □

Now we move to a definition that only makes sense in a market with a numéraire.

Definition. Let P 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.

Suppose that there exists a numéraire portfolio η and let $N = \eta \cdot P$. An *equivalent martingale measure* relative to this numéraire is any probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ such that the *discounted price* processes

$$\left(\frac{P_t}{N_t} \right)_{t \geq 0}$$

is a martingale under $\mathbb{Q}$.

Remark. In many accounts of arbitrage theory, the concept of an equivalent martingale measure has taken centre stage. I believe that its importance has been overstressed. In particular, it is a *numéraire-dependent* concept, unlike that of a martingale deflator. For instance, if there are two assets that both numéraires (for example from the point of view of a British trader, both the euro and the US dollar are numéraires) then one must be very careful to specify which one is the numéraire.

Before proceeding, it might be useful to recall some facts from probability theory.

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 $\mathbb{P} \sim \mathbb{Q}$, iff

$$\{A \in \mathcal{F} : \mathbb{P}(A) = 1\} = \{A \in \mathcal{F} : \mathbb{Q}(A) = 1\}$$

The above definition says that equivalent probability measures have the same almost sure events. Complementarily, equivalent probability measures have the same null sets: that is, $\mathbb{P} \sim \mathbb{Q}$ iff

$$\{A \in \mathcal{F} : \mathbb{P}(A) = 0\} = \{A \in \mathcal{F} : \mathbb{Q}(A) = 0\}.$$

It turns out that equivalent measures can be characterised 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 $\mathbb{P}$-a.s. positive random variable Z such that*

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

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

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

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

As a mnemonic device, note that the Radon–Nikodym derivative satisfies the identity

$$\mathbb{Q}(A) = \int_A \frac{d\mathbb{Q}}{d\mathbb{P}} d\mathbb{P}.$$

Also, note that $\mathbb{P}$ has a density with respect to $\mathbb{Q}$ given by

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

We only need the easy direction of the theorem, that the existence of a positive density implies equivalence, for this course. Here is a proof. The proof of the harder direction is omitted since we do not need it.

Proof. Suppose $\mathbb{P}(Z > 0) = 1$ and that $\mathbb{E}^{\mathbb{P}}(Z) = 1$. Define a set function $\mathbb{Q}$ by

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

Note that $\mathbb{Q}$ is countably additive by the monotone convergence theorem. Also, $\mathbb{Q}(\Omega) = \mathbb{E}^{\mathbb{P}}(Z) = 1$, so $\mathbb{Q}$ is a probability measure. If $\mathbb{P}(A) = 0$, then the event $\{\mathbb{1}_A = 0\}$ is $\mathbb{P}$-almost sure and hence

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}(Z\mathbb{1}_A) = 0.$$

Conversely, if $\mathbb{Q}(A) = 0$ we can conclude that $\{Z\mathbb{1}_A = 0\}$ is $\mathbb{P}$-a.s. by the pigeon-hole principle since $\{Z\mathbb{1}_A \geq 0\}$ is $\mathbb{P}$-a.s. But since $\{Z > 0\}$ is $\mathbb{P}$-a.s., we must conclude that $\{\mathbb{1}_A = 0\}$ is $\mathbb{P}$-a.s., i.e. $\mathbb{P}(A) = 0$. Thus $\mathbb{Q}$ and $\mathbb{P}$ are equivalent. $\square$

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

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

Then $\mathbb{P}$ and $\mathbb{Q}$ are equivalent. We may take their density $Z = \frac{d\mathbb{Q}}{d\mathbb{P}}$ to be

$$Z(1) = \frac{1}{500},\ Z(2) = \frac{999}{500},\ Z(3) = 0.$$

(Since both measures don't 'see' the event $\{3\}$, we can let $Z(3)$ be any value.)

Now, returning to our financial model, we have a result that says, that in a market with a numéraire, the notion of an equivalent martingale measure is morally the same as the notion of a martingale deflator.

Proposition. *Suppose the market has a numéraire, and fix a non-random time horizon $T > 0$. The market model $(P_t)_{0 \le t \le T}$ has an equivalent martingale measure relative to the numéraire if and only if there is a martingale deflator.*

Proof. Let Y be a process such that $\{Y_T > 0\}$ is $\mathbb{P}$-a.s. and such that $Y_T P_T$ is $\mathbb{P}$-integrable. Define a new measure $\mathbb{Q}$ by the density

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{Y_T N_T}{\mathbb{E}^{\mathbb{P}}(Y_T N_T)}.$$

Our analysis turns on the Bayes formula

$$\mathbb{E}^{\mathbb{Q}}\left(\frac{P_T}{N_T}|\mathcal{F}_t\right) = \frac{\mathbb{E}^{\mathbb{P}}\left(P_T Y_T|\mathcal{F}_t\right)}{\mathbb{E}^{\mathbb{P}}\left(N_T Y_T|\mathcal{F}_t\right)}$$

Suppose Y is a martingale deflator. We have

$$\mathbb{E}^{\mathbb{P}}\left(P_T Y_T|\mathcal{F}_t\right) = P_t Y_t.$$

by definition. Also note that

$$Y_t N_t - Y_{t-1}N_{t-1} = \eta_t \cdot (Y_t P_t - Y_{t-1}P_{t-1})$$

and hence YN is a local martingale. However, since YN is non-negative, we know from last section that YN is a true martingale. In particular

$$\mathbb{E}^{\mathbb{P}}\left(N_T Y_T|\mathcal{F}_t\right) = N_t Y_t.$$

By the Bayes formula we have

$$\mathbb{E}^{\mathbb{Q}}\left(\frac{P_T}{N_T}|\mathcal{F}_t\right) = \frac{P_t}{N_t}$$

and hence P/N is a $\mathbb{Q}$-martingale, i.e. $\mathbb{Q}$ is an equivalent martingale measure.

Conversely, suppose $\mathbb{Q}$ is an equivalent martingale measure. Let

$$Z_t = \mathbb{E}^{\mathbb{P}}\left(\frac{d\mathbb{Q}}{d\mathbb{P}}|\mathcal{F}_t\right).$$

Note that Z is a positive $\mathbb{P}$-martingale. Let

$$Y_t = Z_t/N_t.$$

Since the random variable P_T/N_T is $\mathbb{Q}$-integrable by the definition of martingale, we can conclude that $P_T Y_T$ is $\mathbb{P}$-integrable. Furthermore, the process Y is positive and satisfies

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\left(N_T Y_T|\mathcal{F}_t\right) &= \mathbb{E}^{\mathbb{P}}(Z_T|\mathcal{F}_t)\\ &= Z_t\\ &= N_t Y_t.\end{aligned}$$

Hence by the Bayes formula

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left(P_T Y_T | \mathcal{F}_t\right) &= \mathbb{E}^{\mathbb{Q}}\left(\frac{P_T}{N_T} | \mathcal{F}_t\right) \mathbb{E}^{\mathbb{P}}\left(N_T Y_T | \mathcal{F}_t\right) \\
&= \frac{P_t}{N_t}(N_t Y_t) \\
&= P_t Y_t
\end{aligned}$$

so that PY is a $\mathbb{P}$-martingale and hence Y is a martingale deflator. □

Combining the two results of this section, we have the usual formulation of the first fundamental theorem of asset pricing:

Theorem (First Fundamental Theorem of Asset Pricing when there is a numéraire). *Suppose the market has is a numéraire, and fix a non-random time horizon $T > 0$. The market model $(P_t)_{0 \le t \le T}$ has no terminal consumption arbitrage if and only if there exists an equivalent martingale measure relative to the numéraire.*

8. Special numéraires and equivalent martingale measures

In this section we consider two classes of numéraire assets (and their respective equivalent martingale measures) that arise very frequently in applications.

Definition. A (risk-free zero-coupon) bond is an asset such that there exists a non-random time $T > 0$ (called its maturity date) and such that its price at time T is a non-random positive constant (called its face value or principal value). Unless otherwise specified, we shall assume that the face value of a bond is £1. We will denote the time t price of the bond of maturity T by P_t^T for $0 \le t \le T$.

Proposition. *Suppose the market contains a bond. If the market has no arbitrage, then the bond is a numéraire.*

Proof. There are at least two ways to prove this. It is important to understand both methods.
A 'primal' argument. The idea is that if the bond price drops to zero or less, then an investor could lock in a risk-less profit. In particular, if there is no arbitrage, the price must stay positive.

In mathematical notation, let

$$\tau = \inf\{0 \le t \le T : P_t^T \le 0\}.$$

with the convention that $\tau = +\infty$ if $P_t^T > 0$ for all $0 \le t \le T$.

Note that τ is a stopping time. Consider predictable process $H_t = \mathbb{1}_{\{\tau < t \le T\}}$. This corresponds to a portfolio of buying the bond immediately after the price drops to zero or below and hold it until maturity. Note that

$$\begin{aligned}
c_0 &= -H_1^T P_0^T = -\mathbb{1}_{\{\tau = 0\}} P_0^T \ge 0 \\
c_t &= (H_t^T - H_{t+1}^T) P_t^T = -\mathbb{1}_{\{\tau = t\}} P_t^T \ge 0 \text{ for } 1 \le t \le T \\
c_T &= \mathbb{1}_{\{\tau < T\}} P_T^T \ge 0.
\end{aligned}$$

If there is no arbitrage, then $c_t = 0$ a.s. for all $0 \le t \le T$. Hence $\tau = +\infty$ a.s.

A 'dual' argument. Since there is no arbitrage, there exists a martingale deflator Y. Note YP^T is a martingale, and hence

$$P_t^T = \frac{1}{Y_t}\mathbb{E}(Y_T|\mathcal{F}_t)$$

since $P_T^T = 1$ by definition. Since $Y_t > 0$ a.s. for all $t \geq 0$, the conclusion follows. □

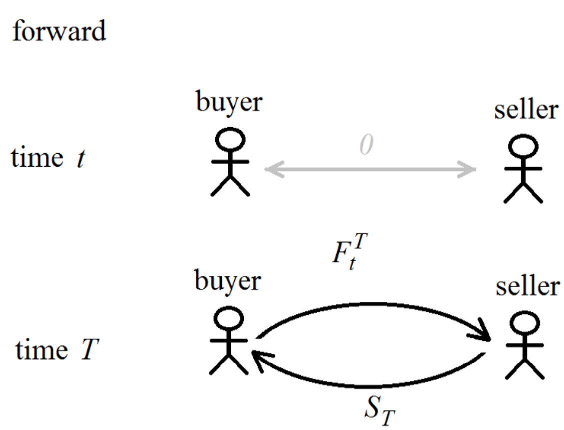
Since we now know no arbitrage implies that bonds are numéraires, we can discuss equivalent martingale measures:

DEFINITION. An equivalent martingale measure with respect to a bond of maturity $T > 0$ is called a T-*forward* measure.

For an application of forward measures, we consider a forward contract:

DEFINITION. A *forward* initiated at time t with maturity date T is a contract where, at time t, no money is exchanged and at time T, an asset with price S_T is swapped for F_t^T units of money, where F_t^T is known at time t. The quantity F_t^T is called the *forward price* of the asset.

See the diagram for the timing of the payment streams in a forward contract.



PROPOSITION. *Consider a market model with a bond of maturity T and forward contract initiated at time t and maturing at time T. If the market has no arbitrage, then*

$$F_t^T = \mathbb{E}^{\mathbb{Q}^T}(S_T|\mathcal{F}_t)$$

where $\mathbb{Q}^T$ is a T-forward measure.

PROOF. Let D_s be the price of the forward contract at time s, for $t \leq s \leq T$. Since there is no arbitrage, we have

$$D_t = P_t^T \ \ \mathbb{E}^{\mathbb{Q}^T}(D_T|\mathcal{F}_t)$$

Noting that $D_t = 0$ and $D_T = S_T - F_t^T$, and that F_t^T is $\mathcal{F}_t$-measurable yields the result, upon rearrangement. □

We now suppose there is a whole family of bonds, indexed by their maturities.

Definition. Consider a market with bonds of all maturities $T \in \{0, 1, 2, \ldots\}$. The *spot interest rate* at time t is

$$r_t = \frac{1}{P^t_{t-1}} - 1.$$

The value of the *bank account* (or *money market account*) is given by $B_0 = 1$ and

$$B_t = \prod_{s=1}^{t}(1 + r_s) \text{ for } t \geq 1.$$

for all t.

Remark. Note that the spot interest rate and bank account processes are previsible.

The connection between bonds and the bank account is the following:

Proposition. *Suppose the arbitrage-free market model has bonds of all maturities. Then there exists a pure-investment (that is, no consumption) strategy η such that*

$$B_t = \eta_t \cdot (P^0_t, P^1_t, P^2_t, \ldots)$$

for all $t \geq 0$.

Proof. Let $\eta_t = B_t \delta^t$, where δ^t is the portfolio of holding exactly one bond of maturity t. That is to say, η is the strategy of investing all of the accumulated wealth in the bond of maturity t during the interval $(t-1, t]$. Note that η is previsible and that

$$\begin{aligned}
\eta_{t+1} \cdot (P^0_t, P^1_t, P^2_t, \ldots) &= B_{t+1} \delta^{t+1} \cdot (P^0_t, P^1_t, P^2_t, \ldots) \\
&= B_t (1 + r_{t+1}) P^{t+1}_t \\
&= B_t \\
&= B_t \delta^t \cdot (P^0_t, P^1_t, P^2_t, \ldots) \\
&= \eta_t \cdot (P^0_t, P^1_t, P^2_t, \ldots).
\end{aligned}$$

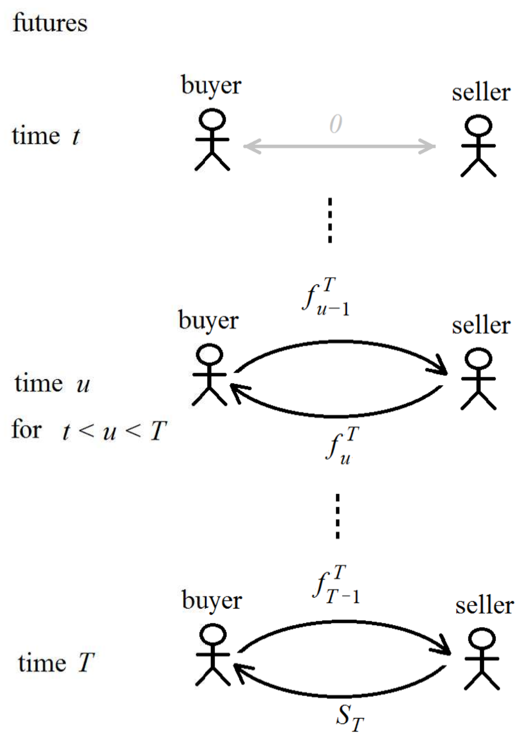
□

Definition. An equivalent martingale measure with respect to the bank account is called a *risk-neutral* measure.

Now we consider a natural application of a risk-neutral measure.

Definition. A *futures contract* initiated at time t with maturity date T is a contract where, at time t, no money is exchanged; at every time $t < u < T$, the sum of f^T_u units of money is swapped for f^T_{u-1} units of money; and finally at time T, the asset with price S_T is swapped for f^T_{T-1} units of money, where f^T_s is known at time s for $t \leq s \leq T$. The quantity f^T_t is called the *futures price* of the asset.

A futures contract is essentially a portfolio of forward contracts of different maturities, with the payout of each forward dependent on the futures prices at the different times before maturity. See the diagram for the timing of the payment streams in a futures contract.



Proposition. *Consider a market model with a bank account and futures contracts for maturity T initiated at each time u for $t \le u < T$. If the market has no arbitrage, then*

$$f_t^T = \mathbb{E}^{\mathbb{Q}}(S_T|\mathcal{F}_t)$$

where $\mathbb{Q}$ is a risk-neutral measure.

Proof. By the 1FTAP says that a market with a dividend paying asset is free of arbitrage if and only if there exists a risk-neutral measure $\mathbb{Q}$, i.e. a measure under which the process M defined by

$$M_t = P_t/B_t + \sum_{s=1}^{t} \delta_s/B_s$$

is a martingale. In our example, we set $\delta_u = f_u^T - f_{u-1}^T$ and $P_u = 0$ for $t \le u \le T$, where $f_T^T = S_T$.

Note

$$\begin{aligned} 0 = \mathbb{E}^{\mathbb{Q}}[M_{u+1} - M_u|\mathcal{F}_u] &= \mathbb{E}^{\mathbb{Q}}[(f_{u+1}^T - f_u^T)/B_{u+1}|\mathcal{F}_u] \\ &= \left(\mathbb{E}^{\mathbb{Q}}[f_{u+1}^T|\mathcal{F}_u] - f_u^T\right)/B_{u+1} \end{aligned}$$

by the $\mathcal{F}_u$-measurability of B_{u+1}. Hence $(f_u^T)_{t \le u \le T}$ is a $\mathbb{Q}$-martingale. □

Remark. Let f_t^T be the futures price of a stock at time t for maturity T. If $T \mapsto f_t^T$ is non-decreasing, the market for that stock is said to be in *contango* at time t. And if $T \mapsto f_t^T$ is non-increasing, the market is said to be in *normal backwardation*.

Forward measures and risk-neutral measures are in general different. In particular, forward prices and futures prices usually disagree. But there is an important example where they agree:

Proposition. *Suppose that the spot interest rate process is not random. A probability measure is a T-forward measure if and only if it is risk-neutral.*

Proof. The fundamental result is that if B is a non-random process then

$$P_t^T = \frac{B_t}{B_T}.$$

This formula can be seen as an instance of the law of one price; see the first example sheet. But here is the quick dual argument: Since both YP^T and YB are martingales we have

$$\begin{aligned} B_t &= \frac{1}{Y_t}\mathbb{E}(B_T Y_T|\mathcal{F}_t) \\ &= \frac{B_T}{Y_t}\mathbb{E}(P_T^T Y_T|\mathcal{F}_t) \\ &= B_T P_t^T. \end{aligned}$$

Now, since B_T is assumed to be a constant, the process $(S_t/B_t)_{0\le t\le T}$ is a martingale with respect to a certain measure if and only if $(S_t/P_t^T)_{0\le t\le T}$ is a martingale with respect to the same measure. □

9. Contingent claim pricing and hedging

The setting of this section is as follows. We find ourselves in a market with prices $(P_t)_{t\ge 0}$. A contingent claim is any cash payment where the size of the payment is contingent on the prices of other assets or any other variable[3] There are two major types of contingent claims that we will study in these notes: European and American.

- **European:** specified by a time horizon $T > 0$ and $\mathcal{F}_T$-measurable random variable ξ_T modelling the payout at the maturity date T.
- **American:** specified by a time horizon $T > 0$ and an adapted process $(\xi_t)_{0\le t\le T}$ where ξ_t models the payout of the claim if the owner of the claim chooses to exercise at time t.

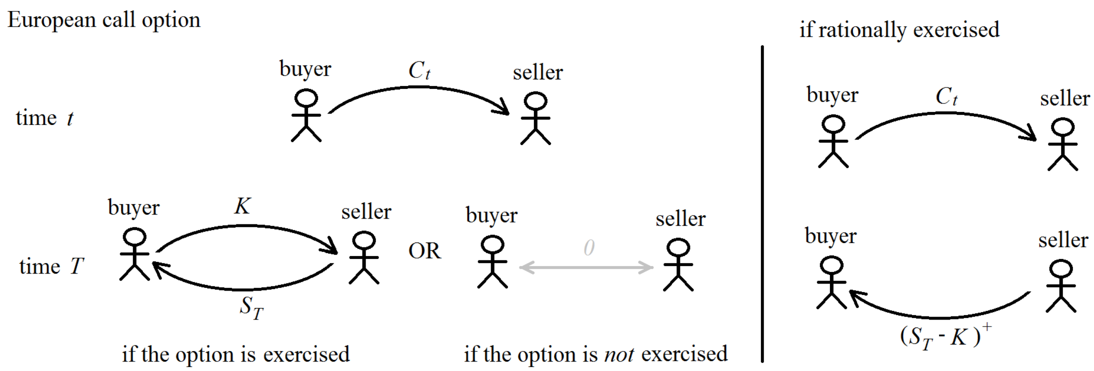
We put ourselves in the shoes of an investment bank that would like to market a new contingent claim. The question are these: what is a 'good' initial price for this claim? How can the seller hedge against the liability of owing the buyer the payout of the claim?

We first consider European options.

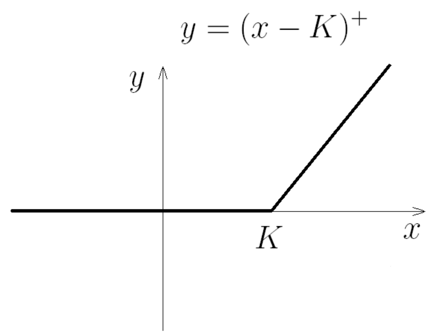
Example (Forward contract). Given a market for a traded asset with prices $(S_t)_{t\ge 0}$, a forward contract initiated at a fixed time t for maturity T is a European claim with payout $\xi_T = S_T - F_t^T$, where F_t^T is the forward price at time t for maturity T. We have discussed this example in the previous section.

[3]... such as the weather. In fact, there exist traded contracts called weather derivatives, marketed to farmers as a hedge against poor growing conditions, energy companies as a hedge against warm winters (and therefore low demand for heating), etc.

Example (Call option). A European call option gives the owner of the option the right, but not the obligation, to buy a given stock at a fixed time T (called the maturity date) at some fixed price K (called the *strike*). See the figure for the timing of the payments, where C_t is the price of the option at time t and S_T is the price of the stock at the maturity date T. There are two cases: If $K \geq S_T$, 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 $K < S_T$, 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_T to the market, realising a profit of $S_T - K$. Hence, the payout of the call option is $\xi_T = (S_T - K)^+$, where $a^+ = \max\{a, 0\}$ as usual. The 'hockey-stick' graph of the function $g(x) = (x - K)^+$ is below.



We will assume that the original market has no arbitrage, since otherwise it is difficult to formulate a reasonable answer to the pricing and hedging questions. Therefore, we will assume that there is at least one martingale deflator.

Proposition. *Consider an arbitrage-free market with prices P. Introduce to this market a European contingent claim with maturity T and payout ξ_T. Suppose that for $t < T$, the price of the claim at time t is ξ_t. If the augmented market with prices $(P_t, \xi_t)_{0 \leq t \leq T}$ has no arbitrage, then*

$$\underline{\xi}_t \leq \xi_t \leq \overline{\xi_t} \text{ for all } 0 \leq t \leq T$$

where

$$\underline{\xi}_t = \text{ess inf}\Big\{\frac{1}{Y_t}\mathbb{E}(Y_T \xi_T | \mathcal{F}_t) : Y \text{ a martingale deflator for the original market such that } \xi_T Y_T \text{ is integrable}\Big\}$$

and $\overline{\xi}_t$ is defined similarly in terms of the essential supremum.

Proof. This is just the 1FTAP. Indeed, if the augmented market there must be a martingale deflator for the augmented market. Such a martingale deflator is necessarily a martingale deflator for the original market. In particular, if Y is a martingale deflator for the augmented market, then $\xi_t = \frac{1}{Y_t}\mathbb{E}(Y_T \xi_T | \mathcal{F}_t)$. That ξ_t is in the interval $[\underline{\xi}_t, \overline{\xi}_t]$ follows directly from the definition of essential infimum and supremum. □

Remark. We pause briefly to discuss the notions of essential supremum and essential infimum.. Given a collection of random variables $(X_k)_{k\in K}$ indexed by an arbitrary (possibly uncountable) set K, a random variable Y is called the *essential supremum* of the collection, denoted

$$Y = \text{ess sup}_{k\in K} X_k$$

iff it satisfies

- $Y \geq X_k$ almost surely for all $k \in K$, and
- if another random variable Z is such that $Z \geq X_k$ almost surely for all $k \in K$, then $Z \geq Y$ almost surely.

The essential infimum of the collection is defined similarly. The proof of the existence of the essential supremum of a family of random variables is on example sheet 2.

Recall, that in contrast, the function $\hat{Y}$ defined by $\hat{Y}(\omega) = \sup_{k\in K} X_k(\omega)$ has the property that $\hat{Y} \geq X_k$ for all $k \in K$ everywhere, and if Z is another function such that $Z \geq X_k$ for all $k \in K$ everywhere, then $Z \geq \hat{Y}$ everywhere.

Here is an example to show that the ordinary notion of supremum is not the correct notion in certain probabilistic settings. Let the set of outcomes Ω be the interval $[0, 1]$, the set of events $\mathcal{F}$ be the Borel sigma-field and the probability measures $\mathbb{P}$ be the Lebesgue measure. Fix a subset $K \subseteq [0, 1]$ and let $X_k = \mathbb{1}_{\{k\}}$. Note that $\hat{Y} = \sup_k X_k = \mathbb{1}_K$. There are a couple of reasons why $\hat{Y}$ is not very useful probabilistically.

Firstly, note that $\hat{Y}$ is a measurable map from Ω to $\mathbb{R}$ if and only if K is a Borel set. Since K was arbitrary, it is not always the case that $\hat{Y}$ is a random variable.

Secondly, even if K is measurable and $\hat{Y}$ is a random variable, it is 'too big'. Indeed, the smaller random variable $Y = 0$ has the property that $Y \geq X_k$ almost surely for all $k \in K$.

Returning to our application for bounding no-arbitrage prices, we don't know a priori whether the set of martingale deflators is countable or not. So the ordinary supremum may not be measurable. Furthermore, we only care about almost sure inequality for each martingale deflator (not inequality simultaneously for all martingale deflators for every outcome[4]), the essential supremum and infimum appearing the statement of the result are appropriate.

The above theorem says that the principle of no-arbitrage usually is not enough to uniquely price a contingent claim. At best, it gives an interval where the no-arbitrage price may lie. However, there is a special class of contingent claims that can be priced uniquely.

Definition. A European contingent claim with payout ξ_T is *replicable* or *attainable* iff there exists an initial wealth x and pure investment strategy H such that $X_T^{x,H} = \xi_T$ almost

[4]Indeed, the inequalities we care about involve conditional expectations, which are only really defined as an equivalence class of random variables that agree on a set of probability one.

surely. (Recall that $X_0 = x, X_t = H_t \cdot P_t$ for $t \geq 1$ and pure investment means $H_{t+1} \cdot P_t = X_t$ for $t \geq 0$.

One of the reasons to single out attainable claims is that there is an unambiguous way to price them according to the no-arbitrage principle:

THEOREM (Characterisation of attainable claims). *Suppose that the market model with n-dimensional price process P has no arbitrage. Let ξ_T be the payout of a European contingent claim with maturity date $T > 0$. The following are equivalent:*

(1) *The claim is atttainable.*
(2) *There exists a unique process $(\xi_t)_{0 \leq t \leq T}$ such that the augmented market (P, ξ) has no arbitrage.*
(3) *There exists a number ξ_0 such that $\mathbb{E}[Y_T \xi_T] = Y_0 \xi_0$ for all martingale deflators (of the original market) such that $Y_T \xi_T$ is integrable.*

PROOF. (1)$\Rightarrow$ (2) This is the law of one price from the first example sheet. Indeed, let ξ_t be the price of the claim and X_t be the price of the replicating portfolio at time t. Let τ be the first time $\xi_t \neq X_t$. One the event $\{\tau < T\}$ do the following: at time τ buy the cheaper one, sell the expensive one, and consume the difference; and at time T unwind both positions for zero cost since $\xi_T = X_T$ by assumption. In notation, let H be the replicating strategy and let consider the strategy $(\tilde{H}, h)$ in the augmented market given by

$$(\tilde{H}_t, h_t) = \mathbb{1}_{\{\tau \leq t-1\}} (H_t, -1) \text{sign}(\xi_\tau - X_\tau) \text{ for } 1 \leq t \leq T$$

and $\tilde{H}_{T+1} = 0$ and $h_{T+1} = 0$. The corresponding consumption stream is

$$c_t = \mathbb{1}_{\{\tau = t\}} |X_t - \xi_t|$$

This strategy would be an arbitrage unless $\mathbb{P}(\tau < T) = 0$. Hence, no arbitrage in the augmented market implies $\xi_t = X_t$ a.s. for all $0 \leq t \leq T$ and hence the price process is uniquely determined by the replicating strategy.

(By the way, the same argument shows that if $X_T^{x,H} = X_T^{y,K}$ almost surely for possibly different strategies H and K initial wealths x and y, then $X_t^{x,H} = X_t^{y,K}$ almost surely for all $0 \leq t \leq T$. In particular, two replicating strategies of an attainable claim yield the same no arbitrage price.) □

PROOF. (2)$\Rightarrow$ (3) Suppose that there is no arbitrage in the augmented market. Then there exists a martingale deflator Y for that market. In particular, this martingale deflator is a martingale deflator for the original market and also for the new asset. That is, $Y\xi$ is a martingale and in particular, $\mathbb{E}(Y_T \xi_T) = Y_0 \xi_0$. Now, assuming that the initial price ξ_0 of the claim is uniquely identified yields the result. □

It remains to prove the implication (3) $\Rightarrow$ (1). We will only prove the $T = 1$ version. Recall that in this setting, the martingale deflator: $Y_0 > 0, Y_1 > 0$ almost surely and $\mathbb{E}(Y_1 P_1) = Y_0 P_0$, can be replaced by the pricing kernel: $\rho > 0$ almost surely and $\mathbb{E}[\rho P_1] = P_0$, by setting $\rho = Y_1 / Y_0$.

In what follows, we will assume that the one-period market with prices P_0, P_1 has no arbitrage. By the fundamental theorem of asset pricing, there exists a pricing kernel. But, given an arbitrary random variable ξ_1, one may worry whether there must exist a pricing kernel ρ such that $\rho\xi_1$ is integrable. Fortunately, it turns out that if there is no arbitrage, there does indeed exist at least one pricing kernel ρ such that $\rho\xi_1$ is integrable. [5]

We will proceed by a series of lemmas.

LEMMA. *Suppose that there exists a number ξ_0 such that $\mathbb{E}(\rho\xi_1) < \xi_0$ for all pricing kernels such that $\rho\xi_1$ is integrable. Then there exists a portfolio $H \in \mathbb{R}^n$ such that*

$$H \cdot P_0 \leq \xi_0 \text{ and } H \cdot P_1 \geq \xi_1 \text{ a.s.}$$

and there is positive probability that at least one of the above inequalities is strict.

PROOF. By assumption, there does not exist a pricing kernel for the augmented market (P, ξ). By the fundamental theorem of asset pricing, there exists an arbitrage in the augmented market, i.e. a portfolio $(\tilde{H}, h) \in \mathbb{R}^{n+1}$ such that

$$\tilde{H} \cdot P_0 + h\xi_0 \leq 0 \text{ and } \tilde{H} \cdot P_1 + h\xi_1 \geq 0 \text{ a.s.}$$

where there is positive probability that at least one of the above inequalities is strict.

Let ρ be a pricing kernel for the original market such that ρP_1 is integrable. Note that we have

$$\begin{aligned} 0 &\leq \mathbb{E}[\rho(\tilde{H} \cdot P_1 + h\xi_1)] \\ &= \tilde{H} \cdot P_0 + h\mathbb{E}(\rho\xi_1) \\ &\leq h[\mathbb{E}(\rho\xi_1) - \xi_0] \end{aligned}$$

Since $\mathbb{E}(\rho\xi_1) - \xi_0 < 0$, we conclude that $h \leq 0$.

We can rule out the case that $h = 0$. For instance, if $h = 0$ we would have $\tilde{H}$ being an arbitrage in the original market - a contradiction.[6] We are left with $h < 0$. This shows that $H = -\tilde{H}/h$ satisfies the conclusion of the lemma. □

LEMMA. *Suppose that there exists a number ξ_0 such that $\mathbb{E}(\rho\xi_1) \leq \xi_0$ for all pricing kernels such that $\rho\xi_1$ is integrable. Then there exists a portfolio $H \in \mathbb{R}^n$ such that*

$$H \cdot P_0 \leq \xi_0 \text{ and } H \cdot P_1 \geq \xi_1 \text{ a.s.}$$

PROOF. For all $k > 0$, we have $\mathbb{E}(\rho\xi_1) < \xi_0 + 1/k$ and hence by the previous lemma there exists a portfolio $H_k \in \mathbb{R}^n$ such that

$$H_k \cdot P_0 \leq \xi_0 + 1/k \text{ and } H_k \cdot P_1 \geq \xi_1 \text{ a.s.}$$

---

[5] To see why, recall from the proof that we let

$$F(h) = e^{h \cdot P_0} + \mathbb{E}[e^{-h \cdot P_1}\zeta]$$

and chose the positive random variable ζ such that F was finite valued, and hence smooth. A good choice is $\zeta = e^{-\|P_1\|^2/2}$. We then showed that no arbitrage implied that there exists a minimiser h^* of F, and hence

$$\rho = \frac{e^{-h^* \cdot P_1}}{e^{h^* \cdot P_0}}\zeta$$

is a pricing kernel.

Now, there is a lot of freedom with a our choice of ζ. Given the random variable ξ_1, we could, for instance choose $\zeta = e^{-\|P_1\|^2/2 - \xi_1^2/2}$. With this choice, it is clear that the resulting pricing kernel ρ is such that $\rho\xi_1$ is bounded, and hence integrable.

[6] Alternatively, note that since $\mathbb{P}(\tilde{H} \cdot P_0 + h\xi_0 = 0 = \tilde{H} \cdot P_1 + h\xi_1) < 1$, we have $h[\mathbb{E}(\rho\xi_1) - \xi_0] > 0$.

Case: $(H_k)_k$ is bounded. In this case, we can pass to a convergent subsequence such that $H_k \to H^*$. Note

$$H^* \cdot P_0 \leq \xi_0 \text{ and } H^* \cdot P_1 \geq \xi_1 \text{ a.s.}$$

as desired.
Case: $(H_k)_k$ is unbounded. Recall from the proof of the 1FTAP the notation

$$\mathcal{U} = \{u \in \mathbb{R}^n : u \cdot P_0 = 0 = u \cdot P_1 \text{ a.s. }\}$$

and

$$\mathcal{V} = \mathcal{U}^\perp.$$

By projecting our given sequence onto $\mathcal{V}$, we can assume that $H_k \in \mathcal{V}$ for all k and that $(H_k)_k$ is unbounded (otherwise, we are back to the previous case). We can pass to a subsequence such that $\|H_k\| \to \infty$ and to a further subsequence such that $\hat{H}_k = H_k/\|H_k\|$ converges to a non-zero[7] is limit $\hat{H} \in \mathcal{V}$. Now dividing the inequalities by $\|H_k\|$ and taking the limit yields

$$\hat{H} \cdot P_0 \leq 0 \leq \hat{H} \cdot P_1 \text{ a.s.}$$

By no arbitrage, we have $\hat{H} \cdot P_0 = 0 = \hat{H} \cdot P_1$ a.s., or in other notation $\hat{H} \in \mathcal{U}$. Since $\hat{H}$ is in $\mathcal{V}$, we have $\hat{H} = 0$, a contradiction. This shows that this second case is impossible. □

PROOF OF (3) ⇒ (1) IN ONE PERIOD. Suppose that there exists a number ξ_0 such that $\mathbb{E}(\rho\xi_1) = \xi_0$ for all pricing kernels such that $\rho\xi_1$ is integrable. Note that

$$\mathbb{E}(\rho\xi_1) \leq \xi_0 \text{ for all } \rho$$

so that there exists a portfolio $H^+ \in \mathbb{R}^n$ such that

$$H^+ \cdot P_0 \leq \xi_0 \text{ and } H^+ \cdot P_1 \geq \xi_1 \text{ a.s.}$$

Similarly,

$$\mathbb{E}[\rho(-\xi_1)] \leq -\xi_0 \text{ for all } \rho$$

so that there exists a portfolio $H^- \in \mathbb{R}^n$ such that

$$H^- \cdot P_0 \leq -\xi_0 \text{ and } H^- \cdot P_1 \geq -\xi_1 \text{ a.s.}$$

Adding this together yields

$$(H^+ + H^-) \cdot P_0 \leq 0 \leq (H^+ + H^-) \cdot P_1 \text{ a.s.}$$

By no arbitrage in the original market, we have

$$(H^+ + H^-) \cdot P_0 = 0 = (H^+ + H^-) \cdot P_1 \text{ a.s.}$$

Hence the portfolios $H = H^+$ and $H = -H^-$ satisfy the desired conclusion. □

[7]We are only considering here the case where $\mathcal{V} \neq \{0\}$. But let's think a bit about the case $\mathcal{V} = \{0\}$. In this case $P_0 = 0 = P_1$ almost surely. So the lemma says that if $\mathbb{E}(\rho\xi_1) \leq \xi_0$ for all positive random variables ρ, then $\xi_1 \leq 0 \leq \xi_0$ almost surely. Can you prove this statement?

Example. (Put-call parity formula) Suppose we start with a market with three assets with prices $(B_t^T, S_t, C_t)_{0\le t\le T}$. The first asset is a bond with maturity date T and unit principal value, so that in particular, $B_T^T = 1$ almost surely. The next asset is a stock. The last asset is a call option on that stock with strike K and maturity T, so that $C_T = (S_T - K)^+$. 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 at a fixed maturity date. If the strike is K and maturity date is T, then a similar argument as we used for the call option, the payout of a put option is $P_T = (K - S_T)^+$.

It turns out that the put option is replicable in the market (B^T, S, C). Indeed, we have the identity

$$\begin{aligned} P_T &= (K - S_T)^+ \\ &= K - S_T + (S_T - K)^+ \\ &= (K, -1, +1) \cdot (B_T^T, S_T, C_T). \end{aligned}$$

Hence $H_t = (K, -1, +1)$ for all $1 \le t \le T$ is a replicating strategy.

Now, suppose we want to assign prices P_t to the put for $0 \le t < T$. The above theorem says there is no arbitrage in the augmented market (B^T, S, C, P) if and only if

$$\boxed{P_t - C_t = K B_t^T - S_t}$$

This is the famous put-call parity formula.

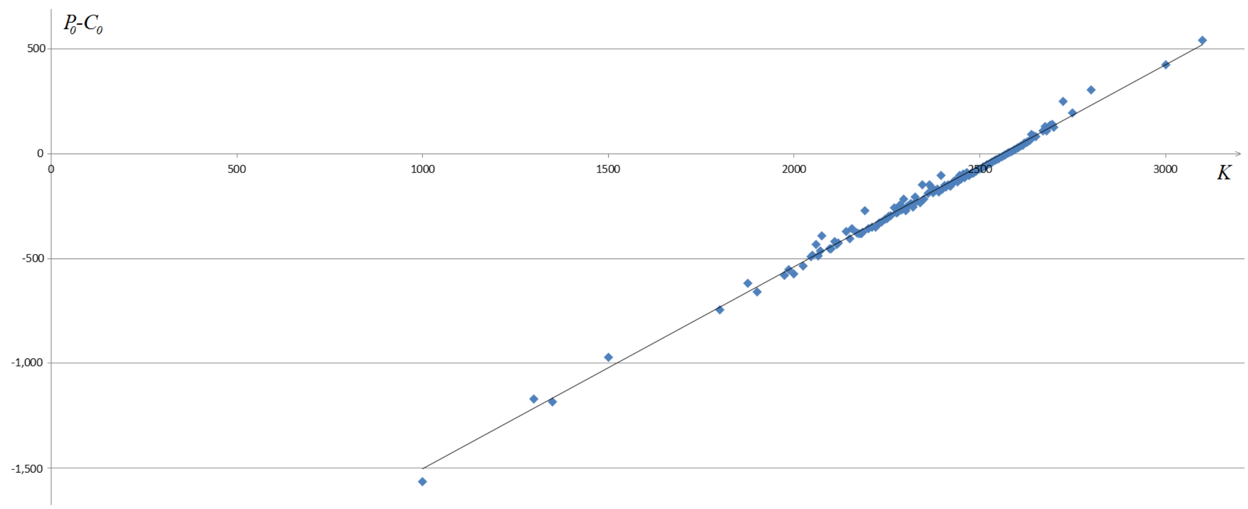


Figure 1. A plot of $P_0 - C_0$ versus K, where $t = 0$ corresponds to 23 October 2017 and and $t = T$ is 17 November 2017, and the underlying asset is the S&P 500 index with $S_0 = 2,573.82$. The price of the calls and puts is taken to be the last traded price on the day (as opposed to the bid or ask price). All data is taken from `https://uk.finance.yahoo.com`.

Since attainable claims have unique no-arbitrage prices, we single out the markets for which every claim is attainable:

Definition. A market is *complete* if and only if every European contingent claim is attainable. A market is *incomplete* otherwise.

In discrete time models complete markets have a lot (probably too much) structure:

Theorem. *If the market model P with n assets is complete, then for each $t \geq 0$ the probability space Ω can be partitioned into no more than n^t $\mathcal{F}_t$-measurable events of positive probability, and in particular, the n-dimensional random vector P_t takes values in a set of at most n^t elements.*

Proof. We will proceed by induction. First suppose $A_1, \ldots, A_k$ are a collection of disjoint $\mathcal{F}_t$-measurable events with $\mathbb{P}(A_i) > 0$ for all i. Claim: the set $\{\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_k}\}$ is linearly independent, and in particular, the dimension of the span of $\{\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_k}\}$ is exactly k. Indeed, we must show that if

$$a_1 \mathbb{1}_{A_1} + \ldots a_k \mathbb{1}_{A_k} = 0 \text{ a.s.}$$

for some constants $a_1, \ldots, a_k$, then $a_1 = \cdots = a_k = 0$. To this end, note that if $i \neq j$ the sets A_i and A_j are disjoint and hence $\mathbb{1}_{A_i} \mathbb{1}_{A_j} = 0$. By multiplying both sides of the equation by $\mathbb{1}_{A_i}$ we get $a_i \mathbb{1}_{A_i} = 0$. But since $\mathbb{P}(A_i) > 0$ it must be the case that $a_i = 0$, proving the claim.

Since the market is complete, each of the $\mathbb{1}_{A_i}$ is replicable. Hence

$$\text{span}\{\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_k}\} \subseteq \{H_t \cdot P_t : H_t \text{ is } \mathcal{F}_{t-1}\text{-meas. }\}$$

Now let $B_1, \ldots, B_N$ be a maximal partition of Ω into disjoint $\mathcal{F}_{t-1}$-measurable sets of positive measure, where by the induction hypothesis $N \leq n^{t-1}$. If a random vector H_t is $\mathcal{F}_{t-1}$-measurable, then it takes exactly one value on each of the B_j's for a total of at most N values $h_1, \ldots, h_N$. Hence

$$\begin{aligned} \{H_t \cdot P_t : H_t \text{ is } \mathcal{F}_{t-1}\text{-meas. }\} &= \{h_1 \cdot P_t \mathbb{1}_{B_1} + \ldots + h_N \cdot P_t \mathbb{1}_{B_N} : h_1, \ldots, h_N \in \mathbb{R}^n\} \\ &= \text{span}\{P_t^i \mathbb{1}_{B_j} : 1 \leq i \leq n, 1 \leq j \leq N\} \end{aligned}$$

and the dimension of the space above is at most nN.

Therefore, we have shown $k \leq nN \leq n^t$, completing the induction. ☐

We can characterise complete markets:

Theorem (Second Fundamental Theorem of Asset Pricing). *An arbitrage-free market model is complete if and only if there exists a unique martingale deflator Y such that $Y_0 = 1$.*

Proof. ('if' direction) Let Y and Y' be martingale deflators with $Y_0 = 1 = Y'_0$. Suppose the market is complete, fix a non-random time $T > 0$ and consider the claim with payout $\xi_T = Y_T - Y'_T$. By completeness, there exists (x, H) such that $X_T^{x,H} = \xi_T$. By completeness, every $\mathcal{F}_T$-measurable random variable is bounded (since it can take at most n^T different values) so both $Y_T \xi_T$ and $Y'_T \xi_T$ are integrable. In particular, we have

$$\mathbb{E}[Y_T \xi_T] = x = \mathbb{E}[Y'_T \xi_T].$$

Subtracting the left and right side of the above equation yields $\mathbb{E}[(Y_T - Y'_T)^2] = 0$ from which the uniqueness of the martingale deflator follows.

('only if' direction) Suppose there is a unique martingale deflator such that $Y_0 = 1$. Then for every contingent claim with payout ξ_T there exists a unique number ξ_0 such that

$\mathbb{E}(Y_T \xi_T) = \xi_0$ for every (that is, the unique) martingale deflator. By the characterisation of attainability, we have ξ_T is attainable, as desired. $\square$

This box summarises the fundamental theorems:

1FTAP:	No arbitrage	$\Leftrightarrow$	Existence of martingale deflator
2FTAP:	Completeness + No arb	$\Leftrightarrow$	Uniqueness of martingale deflator

Finally, we close this section with another useful consequence of completeness:

PROPOSITION. *Suppose the arbitrage-free market model in complete. Then there exists a bank account.*

PROOF. By completeness, bonds of all maturities can be replicated. Hence a bank account can be constructing by holding all the wealth during the period $(t-1, t]$ in the bond with maturity t. $\square$

10. Replication with calls and puts

We consider a market consisting of a bond, a stock with time-T price $S_T \geq 0$, and a family of European calls and puts with strikes in a finite set $\mathcal{K} = \{K_1, \ldots, K_N\} \subseteq (0, \infty)$ all with maturity T.

THEOREM. *Suppose g is piece-wise linear with kinks precisely at the points $\mathcal{K}$. Then the European claim with time T payout $\xi_T = g(S_T)$ can is attainable.*

PROOF. Note that g is differentiable everywhere except $\mathcal{K}$. For every $a \notin \mathcal{K}$, note that the following identity holds

$$g(s) = g(a) + g'(a)(s-a) + \sum_{K \in \mathcal{K}, K < a} \Delta_K (K - s)^+ + \sum_{K \in \mathcal{K}, K > a} \Delta_K (s - K)^+$$

where $\Delta_K = g'(K+) - g'(K-)$. Hence the replicating strategy is to hold $g(a) - ag'(a)$ shares of the bond, to hold $g'(a)$ shares of stock, and holding Δ_K puts of strike $K < a$ and Δ_K calls of strike $K > a$ for all $K \in \mathcal{K}$. $\square$

REMARK. To prove the identity, note that

$$g(x) = g(0) + g'(0)x + + \sum_{K \in \mathcal{K}} \Delta_K (x - K)^+$$

for all $x \geq 0$ and

$$g'(x) = g'(0) + \sum_{K \in \mathcal{K}, K < x} \Delta_K$$

for $x \notin K$, so that

$$\begin{aligned} g(s) - g(a) - g'(a)(s-a) &= \sum_{K \in \mathcal{K}} \Delta_K [(s-K)^+ - (a-K)^+ - (s-a)\mathbb{1}_{\{K<a\}}] \\ &= \sum_{K \in \mathcal{K}, K < a} \Delta_K (K - s)^+ + \sum_{K \in \mathcal{K}, K > a} \Delta_K (s - K)^+. \end{aligned}$$

Remark. In fact, when g is twice continuously differentiable, the continuous analogue of the above identity holds

$$g(s) = g(a) + g'(a)(S_1 - a) + \int_0^a g''(K)(K-s)^+ dK + \int_a^\infty g''(K)(s-K)^+ dK$$

for any $a > 0$. Note that the integrand of the first integral is zero unless $\min\{s,a\} \le K \le a$. Similarly, the integrand of the second integral is zero unless $a \le K \le \max\{s,a\}$. In particular, the ranges of both integrals are bounded intervals on which g'' is assumed continuous, so both integrals are ordinary Riemann integrals.

(One way to prove this identity is to fix s and let $h(a)$ equal the right-hand side. By the standard rules of calculus, we have $h'(a) = 0$ and hence $h(a)$ is a constant. To evaluate that constant, let $a = s$ and note that both integrals vanish since the ranges of integration have zero length.)

Consider the case $g(s) = \log s$. We have the identity

$$\log S_T = \log a + \frac{S_T - a}{a} - \int_0^a \frac{(K-S_T)^+}{K^2} dK - \int_a^\infty \frac{(S_T-K)^+}{K^2} dK.$$

This formula is interpreted to mean that a claim with payout $\xi_T = \log S_T$ can be approximately replicated by trading in calls and puts over a large number of strikes.

Although in reality there do not exist contingent claims with log payouts, market practitioners often think of this log contract as being traded since it can be manufactured (approximately) by calls and puts in a straight-forward manner. This line of thinking has lead to the introduction of variance swap contracts which we will consider in the chapter on continuous time models.

The above result says that given enough call prices, it is possible to replicate any claim with payout of the form $g(S_T)$, assuming g is piece-wise linear (or at least to replicate approximately in the case where g is smooth enough to be approximated by a piece-wise linear function). Assuming we know the initial prices of theses call, we can then calculate the initial cost of the approximate hedging portfolio.

We now come to a simple result observed by Breeden and Litzenberger in 1978. Our setting is a market with prices (P^T, S) where the first asset is a bond of maturity T, let $\mathbb{Q} \sim \mathbb{P}$ be a T-forward measure, so that $\mathbb{E}^{\mathbb{Q}}(S_T) = S_0/P_0^T$. Let $C_T^{T,K} = (S_T - K)^+$ be the payout of a call with strike $K \ge 0$, and let the initial prices be

$$C_0^{T,K} = P_0^T \ \mathbb{E}^{\mathbb{Q}}[(S_T - K)^+].$$

Then for any collection of strikes $\mathcal{K} = \{K_1, \ldots, K_N\}$ the augmented market with prices $(P^T, S, C^{T,K})_{K \in \mathcal{K}}$ has no arbitrage by the easy direction of the first fundamental theorem, since $\mathbb{Q}$ is a forward measure for the augmented market.

Note that the function $K \mapsto C_0^{T,K}$ is convex and therefore has right- and left- derivatives at each point. The following gives meaning to these derivatives:

Proposition (Breeden–Litzenberger formula). *For any $K \ge 0$ we have*

$$\mathbb{Q}(S_T > K) = \frac{1}{P_0^T} D_K^+ C_0^{T,K}$$

and

$$\mathbb{Q}(S_T \geq K) = \frac{1}{P_0^T} D_K^- C_0^{T,K}.$$

If $K \mapsto C_0^{T,K}$ is twice-differentiable then the law of the random variable S_T has a density f_{S_T} under $\mathbb{Q}$ given by

$$f_{S_T}(K) = \frac{1}{P_0^T} D_K^2 C_0^{T,K}.$$

PROOF. Note that

$$\begin{aligned} D^+ C_0(K) &= \lim_{\varepsilon \downarrow 0} \frac{C_0(K+\varepsilon) - C_0(K)}{\varepsilon} \\ &= -P_0^T \lim_{\varepsilon \downarrow 0} \mathbb{E}^{\mathbb{Q}}[g_\varepsilon(S_T - K)] \end{aligned}$$

where

$$g_\varepsilon(x) = \frac{x}{\varepsilon} \mathbb{1}_{[0,\varepsilon)} + \mathbb{1}_{[\varepsilon,\infty)}(x).$$

Note that g_ε is bounded and $g_\varepsilon \to \mathbb{1}_{(0,\infty)}$ pointwise, so the first formula is proven by by the dominated convergence theorem. The formula for the left-derivative is proven similarly.

Finally, if C_0 is twice-differentiable the density is recovered by differentiating once more with respect to K. □

11. Call prices from moment generating functions

Since a portfolio of calls and puts on a stock can essentially replicate any European contingent claim, it is important to have models where the call prices can be computed easily. Unfortunately, there are few models where there exists nice, elementary formulae for the call prices. However, there are many models (especially when we get to continuous time) where the moment generating functions can be computed explicitly, and we will now see that given the moment generating function we can compute call prices by integration:

Consider a market model with a bond of maturity T, a stock with $S_T \geq 0$ almost surely, and let $\mathbb{Q}$ be a fixed T-forward measure. Let

$$C_0^{T,K} = P_0^T \; \mathbb{E}^{\mathbb{Q}}[(S_T - K)^+]$$

for $K > 0$, so that if $C_0^{T,K}$ is the initial price of a call with strike K, then the augmented market has no arbitrage.

For complex θ in the vertical strip

$$\Theta = \{\theta = p + \mathrm{i}q : 0 \leq p \leq 1, q \in \mathbb{R}\}$$

define the moment generating function of the log stock price by

$$M(\theta) = \mathbb{E}^{\mathbb{Q}}(e^{\theta \log S_T} \mathbb{1}_{\{S_T > 0\}}).$$

Note that we have for $\theta = p + \mathrm{i}q \in \Theta$,

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}(|e^{\theta \log S_T} \mathbb{1}_{\{S_T > 0\}}|) &= \mathbb{E}^{\mathbb{Q}}(S_T^p) \\ &\leq \mathbb{E}^{\mathbb{Q}}(S_T)^p \\ &= \left(\frac{S_0}{P_0^T}\right)^p < \infty \end{aligned}$$

by Jensen's inequality, so the moment generating function is well-defined and finite-valued – indeed it is analytic on Θ but we do not use this fact below. The following result shows how to recover call prices from the moment generating function.

THEOREM. *For any $0 < p < 1$ the identity*

$$C_0^{T,K} = S_0 - \frac{K^{1-p} P_0^T}{2\pi} \int_{-\infty}^{\infty} \frac{M(p+ix) e^{-\mathrm{i}x \log K}}{(x - \mathrm{i}p)(x + \mathrm{i}(1-p))} dx$$

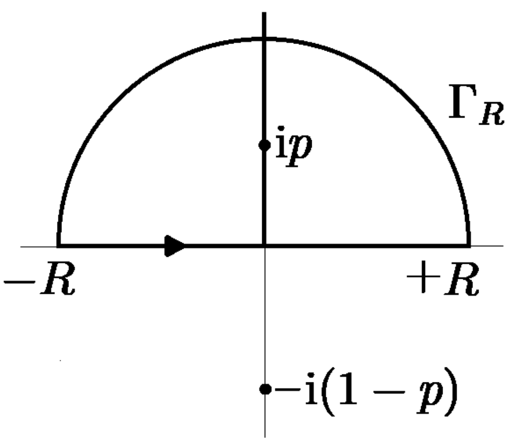
holds.

Essentially, we are inverting the moment generating function via a complex integral. Variants of this procedure are often called a Bromwich, Fourier or Mellin transform. To prove this formula, we begin with a lemma:

LEMMA. *For any $0 < p < 1$ the identity*

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{\mathrm{i}ax}}{(x - \mathrm{i}p)(x + \mathrm{i}(1-p))} dx = \begin{cases} e^{-ap} & \text{if } a \geq 0 \\ e^{a(1-p)} & \text{if } a < 0 \end{cases}$$

holds.



PROOF. This is a standard application of the Cauchy residue theorem. Consider the case $a \geq 0$. Define the semi-circular contour

$$\Gamma_R = \{x + \mathrm{i}0 : -R \leq x \leq R\} \cup \{Re^{\mathrm{i}\phi} : 0 \leq \phi \leq \pi\}$$

in the upper half-plane. Cauchy's theorem

$$\begin{aligned} \int_{\Gamma_R} \frac{e^{\mathrm{i}az}}{(z - \mathrm{i}p)(z + \mathrm{i}(1-p))} dz &= \mathrm{i}2\pi \left. \frac{e^{\mathrm{i}az}}{z + \mathrm{i}(1-p)} \right|_{z=\mathrm{i}p} \\ &= 2\pi e^{-ap} \end{aligned}$$

since the integrand is meromorphic with a simple pole at $z = \mathrm{i}p$ inside the contour, and the contour integral is evaluated in the anticlockwise sense.

On the other hand,

$$\int_{\Gamma_R} \frac{e^{\mathrm{i}az}}{(z - \mathrm{i}p)(z + \mathrm{i}(1-p))} dz = \int_{-R}^{R} \frac{e^{\mathrm{i}ax}}{(x - \mathrm{i}p)(x + \mathrm{i}(1-p))} dx + \int_0^{\pi} \frac{\mathrm{i}Re^{-aR\sin\phi} e^{\mathrm{i}(aR\cos\phi + \phi)}}{(Re^{\mathrm{i}\phi} - \mathrm{i}p)(Re^{\mathrm{i}\phi} + \mathrm{i}(1-p))} d\phi$$

and the second integral vanishes as $R \to \infty$ since $a \geq 0$.

The case $a < 0$ is handled in exactly the same way; just integrate around a semi-circular contour in the lower half-plane enclosing the other pole at $-\mathrm{i}(1-p)$. □

Proof of theorem. From the lemma, we have the identity

$$\begin{aligned}(S_T - K)^+ S_T - S_T \wedge K \\ &= S_T - \frac{K^{1-p}}{2\pi}\int_{-\infty}^{\infty} \frac{e^{p\log S_T + ix\log(S_1/K)}\mathbb{1}_{\{S_T>0\}}}{(x-\mathrm{i}p)(x+\mathrm{i}(1-p))}dx.\end{aligned}$$

Now multiply by P_0^T and compute expectations. The result follows upon interchanging expectation and integration on the right-hand side. This is justified by Fubini's theorem since

$$\int_{-\infty}^{\infty} \mathbb{E}^{\mathbb{Q}}\left|\frac{e^{p\log S_T + ix\log(S_1/K)}}{(x-\mathrm{i}p)(x+\mathrm{i}(1-p))}\right| dx = M(p)\int_{-\infty}^{\infty}\frac{dx}{\sqrt{(x^2+p^2)(x^2+(1-p)^2)}} < \infty$$

□

12. Super-replication of 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. 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.

The payout of an American claim is specified by two ingredients:

- a maturity date $T > 0$,
- an adapted process $(\xi_t)_{0\le t\le T}$.

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.

Now, if an American claim matures at $T > 0$ and is specified by the payout process $(\xi_t)_{0\le t\le T}$, 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:

The market model $P = (P_t)_{0\le t\le T}$ is complete.

Let $Y = (Y_t)_{0\le t\le T}$ be the unique martingale deflator such that $Y_0 = 1$.

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

$$\sup_{\tau\le T}\mathbb{E}\left(Y_\tau \xi_\tau\right)$$

to be sure that he can hedge the option, where the supremum is taken over the set of stopping times smaller than or equal to T. Indeed, this is the case.

Theorem. *Suppose that the adapted process $(\xi_t)_{0\le t\le T}$ specifies the payout of an American claim maturing at $T > 0$.*

There exists a self-financing pure-investment trading strategy H such that

- $X_t \ge \xi_t$ *for all* $0 \le t \le T$,

- $X_{\tau^*} = \xi_{\tau^*}$ *for some stopping time* τ^*, *and*
- $X_0 = \sup_{\tau \le T} \mathbb{E}\,(Y_\tau \xi_\tau)$.

where $X_t = H_t \cdot P_t = H_{t+1} \cdot P_t$.

Remark. The strategy H dominates the payout of the American claim at all times, but is conservative in the sense that it exactly replicates the optimally exercised claim.

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

*****

We will need a result of general interest:

Theorem (Doob decomposition theorem). *Let U be a discrete-time supermartingale. Then there is a unique decomposition*

$$U_t = U_0 + M_t - A_t$$

where M is a martingale and A is a predictable non-decreasing process with $M_0 = A_0 = 0$.

Proof. Let $M_0 = 0 = A_0$ and define

$$\begin{aligned} M_{t+1} &= M_t + U_{t+1} - \mathbb{E}(U_{t+1}|\mathcal{F}_t) \\ A_{t+1} &= A_t + U_t - \mathbb{E}(U_{t+1}|\mathcal{F}_t) \end{aligned}$$

for $t \ge 0$. Since U is assumed to be supermartingale, and hence integrable, the processes M and A are integrable. It is straightforward to check that M is a martingale, and since U is a supermartingale, that A is non-decreasing. Also by induction, we see that A_{t+1} is $\mathcal{F}_t$-measurable.

Summing up,

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

To show uniqueness, assume that $U_t = U_0 + M_t - A_t = U_0 + M'_t - A'_t$. Then $M - M'$ is a predictable discrete-time martingale, that is, a constant. $\square$

Now we introduce the key concept in optimal stopping theory:

Definition. Let $(Z_t)_{0 \le t \le T}$ be a given integrable adapted discrete-time process. Define an adapted process $(U_t)_{0 \le t \le T}$ by the recursion

$$\begin{aligned} U_T &= Z_T \\ U_t &= \max\{Z_t, \mathbb{E}(U_{t+1}|\mathcal{F}_t)\} \text{ for } 0 \le t \le T-1. \end{aligned}$$

The process $(U_t)_{0 \le t \le T}$ is called the *Snell envelope* of $(Z_t)_{0 \le t \le T}$.

Remark. The Snell envelope clearly satisfies both

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

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 Z will be the process $Y\xi$, where Y is the martingale deflator and ξ is the process specifying the payout of the American claim.

Theorem. *Let $(Z_t)_{0\leq t\leq T}$ be an integrable adapted process, let $(U_t)_{0\leq t\leq T}$ be its Snell envelope with Doob decomposition $U_t = U_0 + M_t - A_t$. Let $A_{T+1} = +\infty$ and*

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

Then τ^* is a stopping time and

$$U_{\tau^*} = U_0 + M_{\tau^*} = Z_{\tau^*}.$$

Proof. That τ^* is a stopping time follows from the fact that the non-decreasing process $(A_t)_{0\leq t\leq T+1}$ is predictable.

Now note that

$$\mathbb{E}(U_{t+1}|\mathcal{F}_t) = \mathbb{E}(U_0 + M_{t+1} - A_{t+1}|\mathcal{F}_t) = U_0 + M_t - A_{t+1}$$

since M is a martingale and A is predictable so that by the definition of Snell envelope

$$U_0 + M_t - A_t = \max\{Z_t, U_0 + M_t - A_{t+1}\}.$$

Note that $A_{\tau^*} = 0$ and hence

$$U_0 + M_{\tau^*} = \max\{Z_{\tau^*}, U_0 + M_{\tau^*} - A_{\tau^*+1}\}.$$

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

$$U_{\tau^*} = U_0 + M_{\tau^*} = Z_{\tau^*}.$$

□

Theorem. *Let Z be an adapted integrable process and let U be its Snell envelope. Then*

$$U_0 = \sup\{\mathbb{E}(Z_\tau) : \text{ stopping time } 0 \leq \tau \leq T\}.$$

Proof. Since U is a supermartingale,

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

for any stopping time τ by the optional sampling theorem. (See example sheet 1.) But since $U_t \geq Z_t$ by construction,

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

for any stopping time τ. But letting $\tau^* = \min\{t \in \{0, \ldots, T\} : A_{t+1} > 0\}$ where $U = U_0 + M - A$ is the Doob decomposition of U, we have

$$U_0 = U_0 + \mathbb{E}(M_{\tau^*}) = \mathbb{E}(Z_{\tau^*}).$$

again by the optional sampling theorem and the previous result. □

Remark. By a similar argument, one can show that

$$U_t = \text{ess sup}\{\mathbb{E}(Z_\tau|\mathcal{F}_t) : \text{ stopping time } t \le \tau \le T\}.$$

for all $0 \le t \le T$. This formula allows us to define the Snell envelope for the infinite horizon case $T = \infty$ and also in the continuous time case.

Definition. If Z is an integrable adapted process, a stopping time σ such that $\mathbb{E}(Z_\sigma) = \sup_{0\le\tau\le T} \mathbb{E}(Z_\tau)$ is called an *optimal stopping time*. Obviously the stopping time τ^* defined above is an optimal stopping time. Example sheet 2 shows how to find another one.

*****

Returning to finance, let $(\xi_t)_{0\le t\le T}$ be the process specifying the payout of an American option, let Y the unique martingale deflator with $Y_0 = 0$ and let $(U_t)_{0\le t\le T}$ be the Snell envelope of $Y\xi$ with Doob decomposition $U = U_0 + M - A$.

We now will use the assumption that the market is complete: let H be strategy such that $H_T \cdot P_T = (U_0 + M_T)/Y_T$. Setting $X_t = H_t \cdot P_t$ note that XY is a martingale since it is a local martingale from before, and since the market is complete, it is also bounded. By the martingale property, we have

$$X_t Y_t = U_0 + M_t$$

for all $0 \le t \le T$. In particular,

- $X_t = (U_0 + M_t)/Y_t \ge U_t/Y_t \ge \xi_t$ for all $0 \le t \le T$,
- $X_{\tau^*} = (U_0 + M_{\tau^*})/Y_{\tau^*} = U_{\tau^*}/Y_{\tau^*} = \xi_{\tau^*}$, and
- $X_0 = U_0 = \sup_{\tau\le T} \mathbb{E}(Y_\tau \xi_\tau)$,

completing the proof of the theorem.

13. A dual approach to optimal stopping

The final result in discrete time is the following dual approach to optimal stopping:

Theorem. *Let $(Z_t)_{0\le t\le T}$ be a discrete-time, integrable adapted process. Then*

$$\sup_{\tau\le T} \mathbb{E}(Z_\tau) = \inf_M \mathbb{E}[\max_{0\le t\le T}(Z_t - M_t)]$$

where the supremum on the left-hand side is taken over stopping times τ and the infimum on the right-hand side is taken over martingales M with $M_0 = 0$.

Proof. Let M be a martingale with $M_0 = 0$. By the optional stopping theorem we have

$$\begin{aligned}\sup_{\tau\le T} \mathbb{E}(Z_\tau) &= \sup_{\tau\le T} \mathbb{E}(Z_\tau - M_\tau) \\ &\le \mathbb{E}[\max_{0\le t\le T}(Z_t - M_t)]\end{aligned}$$

Since the inequality holds for any martingale M, it also holds when we take the infimum of the right-hand side over M.

For the reverse inequality, let U be the Snell envelope of Z and let $U = U_0 + M^* - A$ be its Doob decomposition. Since $U_t \leq Z_t$ for all t we have

$$\begin{aligned}
\inf_M \mathbb{E}[\max_{0\leq t\leq T}(Z_t - M_t)] &\leq \mathbb{E}[\max_{0\leq t\leq T}(Z_t - M_t^*)] \\
&\leq \mathbb{E}[\max_{0\leq t\leq T}(U_t - M_t^*)] \\
&= \mathbb{E}[\max_{0\leq t\leq T}(U_0 - A_t)] \\
&= U_0
\end{aligned}$$

The conclusion follows from the last section where we proved that

$$U_0 = \sup_{\tau\leq T}\mathbb{E}(Z_\tau)$$

$\square$

CHAPTER 2

Brownian motion and stochastic calculus

Despite the elegance of discrete-time financial theory, there is at least one glaring problem: explicit computations are difficult. For instance, the fundamental theorems are stated in terms of state price densities, but it is very difficult to classify them except in a few simple examples. The continuous-time theory has the convenient feature that explicit formulae are easy to find–indeed, one of our first results will be the general formula for a state price density in a continuous-time market model.

Before we can describe the continuous-time financial theory, we need to first learn about stochastic integration. Recall that in discrete time, the self-financing condition and budget constraint imply that for the wealth process X corresponding to a pure investment strategy H satisfies

$$X_t Y_t = X_0 Y_0 + \sum_{s=1}^{t} H_s \cdot (Y_s P_s - Y_{s-1} P_{s-1})$$

Recall that when Y is a martingale deflator, the process $M = YP$ is a martingale and the process XY is a local martingale.

The continuous time analogue ought to be something like

$$X_t = X_0 + \int_0^t H_s \cdot dM_s$$

What does the integral on the right mean?

If we assume that the sample paths $t \mapsto M_t$ are differentiable, we could interpret the integral as the Lebesgue integral

$$\int_0^t H_s \cdot \frac{dM_s}{ds} ds.$$

Unfortunately, it turns out that life is not that simple. Now, a theorem of stochastic calculus says that a continuous martingale with everywhere differentiable sample paths is necessarily constant. So if we insist that our price processes have differentiable sample paths, we will have a very boring theory.

This chapter is concerned with an integration theory where we use the martingale property, rather than the differentiability of the sample paths, as the key ingredient. This theory is nice, and indeed something like the fundamental theorem of calculus holds. This means we can do explicit computations.

The most basic example of a continuous martingale is Brownian motion:

1. Brownian motion

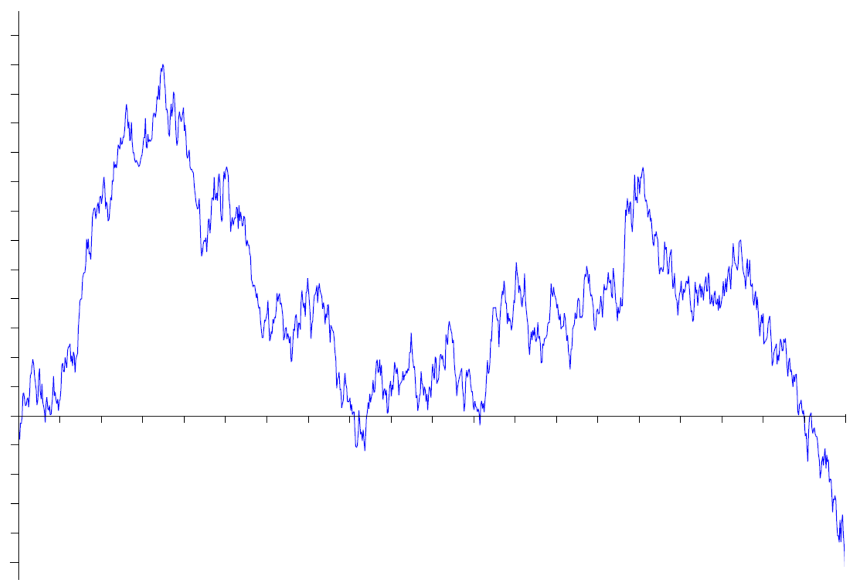
In this section, we introduce one of the most fundamental continuous-time stochastic processes, Brownian motion. As hinted above, our primary interest in this process is that it will be the building block for all of the continuous-time market models studied in these lectures.

Definition. A *Brownian motion* $W = (W_t)_{t \geq 0}$ is a collection of random variables such that

- $W_0(\omega) = 0$ for all $\omega \in \Omega$,
- for all $0 \leq t_0 < t_1 < ... < t_n$ 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 for 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 $(W_t)_{t \geq 0}$ 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 1923. 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:



It will often be useful to talk about a Brownian motion in the context of a filtered probability space:

Definition. A *Brownian motion in a filtration* $(\mathcal{F}_t)_{t \geq 0}$ $W = (W_t)_{t \geq 0}$ is a Brownian such that W_s is $\mathcal{F}_s$-measurable and the increment $W_t - W_s$ is independent of $\mathcal{F}_s$ for all $0 \leq s \leq t$.

Remark. It is a little exercise in measure theory to show that a Brownian motion is automatically a Brownian motion in its natural filtration.

Here are some useful properties of Brownian motion related to the martingale property:

Proposition. *Let W be a Brownian motion in a filtration $\mathcal{F}$. Let $Q_t = W_t^2 - t$ and $Z_t = e^{\alpha W_t - \alpha^2 t/2}$ where $\alpha \in \mathbb{R}$ is constant. Then the process W, Q and Z are all martingales.*

PROOF. Since W_t is $N(0,t)$, we have that W_t, Q_t and Z_t are integrable. Also for $0 \le s \le t$ we have by the measurability of W_s and the independence of $W_t - W_s$ that

$$\begin{aligned}
\mathbb{E}[W_t|\mathcal{F}_s] &= \mathbb{E}[W_t - W_s + W_s|\mathcal{F}_s] \\
&= \mathbb{E}[W_t - W_s] + W_s \\
&= W_s.
\end{aligned}$$

Note that

$$\mathbb{E}[(W_t - W_s)W_s|\mathcal{F}_s] = W_s\mathbb{E}[(W_t - W_s)|\mathcal{F}_s] = 0$$

so

$$\begin{aligned}
\mathbb{E}[W_t^2 - t|\mathcal{F}_s] &= \mathbb{E}[(W_t - W_s + W_s)^2|\mathcal{F}_s] - t \\
&= \mathbb{E}[(W_t - W_s)^2] + W_s^2 - t \\
&= W_s^2 - s
\end{aligned}$$

and finally by the same idea...

$$\begin{aligned}
\mathbb{E}[Z_t] &= Z_s\mathbb{E}[e^{\alpha(W_t - W_s) - \alpha^2(t-s)/2}|\mathcal{F}_s] \\
&= Z_s.
\end{aligned}$$

□

2. Itô stochastic integration

We now have sufficient motivation to construct a stochastic integral with respect to a continuous local martingale. 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, see the lecture notes on my webpage.

For the record, we will assume henceforth that the filtration 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. These are technical assumptions that ensure the existence of stopping times with the right properties.

2.1. Quadratic variation.

THEOREM. *Let M be a continuous local martingale. There exists a finite-valued non-decreasing adapted process A such that A_0 and for all $t \ge 0$ we have*

$$\sum_{i=1}^{n} (M_{s_i^n} - M_{s_{i-1}^n})^2 \to A_t$$

where the convergence is in probability, and $s_i^n = it/n$.

DEFINITION. Given the continuous local martingale M, the non-decreasing process A is called the quadratic variation of M and is denoted $A = \langle M \rangle$.

THEOREM. *The quadratic variation of a Brownian motion W is given by $\langle W \rangle_t = t$ a.s. for all $t \ge 0$.*

Proof. By definition, the increments of Brownian motion are Gaussian random variables so that

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

and

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

for every $0 \le s \le t$. Fix t and n and let $s_i = it/n$.

$$\mathbb{E}\left[\sum_{i=1}^n (W_{s_i} - W_{s_{i-1}})^2\right] = \sum_{i=1}^n (s_i - s_{i-1}) = s_n - s_0 = t$$

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

$$\mathrm{Var}\left[\sum_{i=1}^n (W_{s_i} - W_{s_{i-1}})^2\right] = 2\sum_{i=1}^n (s_i - s_{i-1})^2 = 2t^2/n.$$

Chebychev's inequality implies

$$\mathbb{P}\left(\left|\sum_{i=1}^n (W_{s_i} - W_{s_{i-1}})^2 - t\right| \ge \epsilon\right) \le \frac{2t^2}{n\epsilon^2} \to 0.$$

$\square$

Remark. For comparison, consider a continuously differentiable function $f : [0, 1] \to \mathbb{R}$. Recall that for such functions there exists a constant $C > 0$ such that $|f(t) - f(s)| \le C|t - s|$ for all $s, t \in [0, 1]$. Hence we have

$$\begin{aligned}\sum_{i=1}^n [f(s_i) - f(s_{i-1})]^2 &\le \sum_{i=1}^n C^2 t^2/n^2 \\ &= C^2 t^2/n \to 0\end{aligned}$$

Since the quadratic variation of a Brownian motion is *positive*, we can conclude that the typical Brownian sample path is not a continuously differentiable function of time.

Definition. The *previsible 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 $0 \le s < t$ and A is $\mathcal{F}_s$-measurable.

A *previsible* process α is a map $\alpha : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ that is $\mathcal{P}$-measurable. Equivalently, a previsible process α is the (t, ω) pointwise limit of a sequence of simple processes $(\alpha^n)_n$ of the form

$$\alpha_t^n(\omega) = \sum_{i=1}^n \mathbb{1}_{(s_{i-1}, s_i]}(t) a_i(\omega)$$

where a_i is bounded and $\mathcal{F}_{s_{i-1}}$-measurable for some non-random $0 \le s_0 < s_1 < ... < s_n < \infty$.

Remark. A continuous adapted process is previsible.

Theorem. *Let M be a continuous local martingale, and let α be a previsible processes such that*

$$\int_0^t \alpha_s^2 d\langle M\rangle_s < \infty \text{ a.s. for all } t \ge 0$$

Then there exists a continuous local martingale X such that

$$\sum_{i=1}^{n} \alpha^n_{s^n_i}(M_{s^n_i} - M_{s^n_{i-1}}) \to X_t$$

in a sense which we don't make precise but involves a sequence of stopping times τ_n localising M and a sequence of simple previsible processes α^n converging to α with respect to a certain norm.

Definition. Given the continuous local martingale M be and the previsible process α, the continuous local martingale X is called the stochastic integral of α with respect to M and is denoted

$$X_t = \int_0^t \alpha_s dM_s$$

Remark. The inspired idea of the above definition of the integral is that it compensates for the roughness of a *typical* sample path of M by using instead the many cancellations that occur *on average* from the uncorrelated increments.

Theorem. *Let M, M^1 and M^2 be continuous local martingales and α, α^1 and α^2 previsible processes. Suppose the integrability conditions hold to ensure the existence of the following stochastic integrals. Then we have*

- $\left\langle \int_0^{\cdot} \alpha_s dM_s \right\rangle_t = \int_0^t \alpha_s^2 d\langle M \rangle_s$.
- $\int_0^t (c\alpha_s) dM_s = \int_0^t \alpha_s d(cM_s) = c\int_0^t \alpha_s dM_s$ *where $c \in \mathbb{R}$ is a constant*
- $\int_0^t (\alpha_s^1 + \alpha_s^2) dM_s = \int_0^t \alpha_s^1 dM_s + \int_0^t \alpha_s^2 dM_s$
- $\int_0^t \alpha_s d(M_s^1 + M_s^2) = \int_0^t \alpha_s dM_s^1 + \int_0^t \alpha_s dM_s^2$
- $\int_0^t \beta_s d\left(\int_0^s \alpha_u dM_u\right) = \int_0^t (\beta_s \alpha_s) dM_s$

We end this section with a criterion for knowing when a local martingale is a true martingale:

Proposition. *Let M be a continuous local martingale. If $\mathbb{E}(\langle M \rangle_t) < \infty$ for all $t \geq 0$ then M is a martingale, and*

$$\mathbb{E}(M_t^2) = M_0^2 + \mathbb{E}(\langle M \rangle_t).$$

In particular if α is previsible and

$$\mathbb{E}\left(\int_0^t \alpha_s^2 ds\right) < \infty \text{ for all } t \geq 0$$

then the stochastic integral $M_t = \int_0^t \alpha_s dW_s$ is a martingale (where W is a Brownian motion).

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.

3. Itô's processes and quadratic variation

In the last section, we sketched very quickly the constructed of a stochastic integral with respect to a continuous local martingale. What makes the Itô stochastic integral useful is that there is a corresponding stochastic calculus. To describe it, we need a few more definitions.

As before, let $(W_t)_{t\geq 0}$ be a scalar Brownian motion in a filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying our usual conditions.

We can use our stochastic integration theory to define a useful class of stochastic process:

Definition. An *Itô process* X is an adapted process of the form

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

where X_0 is a fixed real number and $(\alpha_t)_{t\geq 0}$ and $(\beta_t)_{t\geq 0}$ be previsible real-valued processes such that

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

almost surely for all $t \geq 0$. For such an Itô process, we use the differential notation

$$dX_t = \alpha_t dW_t + \beta_t dt$$

or even

$$dX = \alpha\ dW + \beta\ dt$$

as short hand for the integral notation. (The sample paths of the Brownian motion are nowhere differentiable, so the notation dW_t is only *formal*, and can only be interpreted via the stochastic integration theory.)

Note that the two integrals appearing the above definition have different meanings: the first as a stochastic integral and the second as a pathwise Lebesgue integral.

We now introduce a notion which helps with computations involving Itô's formula.

Theorem. *Let X be an Itô process as above. Then*

$$\sum_{i=1}^n (X_{s_i^n} - X_{s_{i-1}^n})^2 \to \int_0^t \alpha_s^2 ds$$

for each $t \geq 0$, where the limit is in probability and $s_i^n = it/n$.

Definition. The quadratic variation of the Itô process X with decomposition

$$dX_t = \alpha_t dW_t + \beta_t dt$$

is given by

$$\langle X \rangle_t = \int_0^t \alpha_s^2 ds$$

or in differential notation

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

4. Itô's formula

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

THEOREM (Itô's formula, scalar version). *Let X be an Itô process. If $f : \mathbb{R} \to \mathbb{R}$ twice continuously differentiable, then $f(X)$ is an Itô process with*

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

Or equivalently, if X is of the form

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

then

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)\alpha_s dW_s + \int_0^t \left[f'(X_s)\beta_s + \frac{1}{2}f''(X_s)\alpha_s^2\right] ds.$$

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

$$\mathbb{E}\left(\int_0^t W_s^2 ds\right) = \int_0^t s\ ds = t^2/2 < \infty,$$

the local martingale X given by

$$X_t = \int_0^t W_s dW_s$$

is actually a *true* martingale. (Remember that we also verified, directly from the definition of Brownian motion, that the process $Q_t = W_t^2 - t$ is a martingale.)

EXAMPLE. Consider the Itô process given by

$$X_t = X_0 + \int_0^t \alpha_s W_s + \int_0^t \beta_s ds$$

for some predictable processes α, β. Letting

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

we would like to show that the process $(Y_t)_{t\geq 0}$ 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,

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

So Itô's formula says:

$$df(X_t) = f'(X_t)dX_t + \frac{1}{2}f''(X_t)d\langle X\rangle_t$$
$$\Rightarrow dY_t = Y_t[(\beta_t + \alpha_t^2/2)dt + \alpha_t\ dW_t]$$

Note that

$$Y \text{ is a local martingale } \Leftrightarrow \beta = -\alpha^2/2.$$

IDEA OF PROOF OF ITÔ'S FORMULA. Fix a partition of $[0, t]$. By telescoping a sum and consider the following second order Taylor approximation we have the following:

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

□

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

An n-dimensional Itô process $(X_t)_{t \geq 0}$ defined by

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

interpreted component-wise as

$$X_t^{(i)} = X_0^{(i)} + \int_0^t \sum_{k=1}^{m} \alpha_s^{(i,k)} dW_s^{(k)} + \int_0^t \beta_s^{(i)} ds$$

where $(W_t)_{t \geq 0}$ is an m-dimensional Brownian motion so that $W^{(1)}, \ldots, W^{(m)}$, are independent scalar Brownian motions, and the previsible process $(\alpha_t)_{t \geq 0}$ is valued in the space of $n \times m$ matrices, and the predictable process $(\beta_t)_{t \geq 0}$ is valued in $\mathbb{R}^n$. We insist that

$$\int_0^t \sum_{i=1}^{n} \sum_{k=1}^{m} (\alpha_s^{(i,k)})^2 ds < \infty \text{ and } \int_0^t \sum_{i=1}^{n} |\beta_s^{(i)}| ds < \infty$$

almost surely for all $t \geq 0$ so that all of the integrals are defined. The aim of this section is to give a formula for the Itô decomposition of $f(t, X_t)$.

Now in the scalar case we needed a notion of quadratic variation $(dX_t)^2 = d\langle X \rangle_t$. In the multi-dimensional case, we now introduce the notion of quadratic co-variation $(dX_t^{(i)})(dX_t^{(j)}) = d\langle X^{(i)}, X^{(j)} \rangle_t$.

DEFINITION. The *quadratic co-variation* of $X^{(i)}$ and $X^{(j)}$ two Itô processes is defined by

$$\langle X^{(i)}, X^{(j)} \rangle_t = \frac{1}{2} \big(\langle X^{(i)} + X^{(j)} \rangle_t - \langle X^{(i)} \rangle_t - \langle X^{(j)} \rangle_t \big)$$

Theorem. *Let X be a multi-dimensional Itô process as above. Then*

$$\langle X^{(i)}, X^{(j)} \rangle_t = \lim_n \sum_{k=1}^{s} (X^{(i)}_{s_k} - X^{(i)}_{s_{k-1}})(X^{(j)}_{s_k} - X^{(j)}_{s_{k-1}})$$
$$= \int_0^t \sum_{k=1}^{m} \alpha_s^{(i,k)} \alpha_s^{(j,k)} ds.$$

for each $t \geq 0$, where $s_k = kt/n$ and where the limit is in probability.

The following multiplication table might help you remember how to compute quadratic co-variation, where W and $W^\perp$ denote independent Brownian motions:

$(dt)^2 = 0$	$(dt)(dW_t) = 0$
$(dW_t)^2 = dt$	$(dW_t)(dW_t^\perp) = 0$

Now we are ready for the statement of the theorem:

Theorem (Itô's formula, multi-dimensional 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*

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

5. Girsanov's theorem

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

Recall that an equivalent measures is related to a positive random variable via the Radon–Nikodym theorem. Indeed, let $(\Omega, \mathcal{F}, \mathbb{P})$ be our probability space and let $\mathbb{Q}$ be equivalent to $\mathbb{P}$. Then, by the Radon–Nikodym theorem there exists a density

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

such that $Z > 0$ has unit $\mathbb{P}$-expectation. Conversely, if $Z > 0$ and $\mathbb{E}^{\mathbb{P}}(Z) = 1$, we can define an equivalent measure $\mathbb{Q}$ with density Z.

Motivated by above discussion, we aim to understand how martingales arise within the context of the Itô stochastic integration theory. Consider the stochastic process $(Z_t)_{t \geq 0}$ given by

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

where $(W_t)_{t\geq 0}$ is a m-dimensional Brownian motion and $(\alpha_t)_{t\geq 0}$ is a m-dimensional predictable process with $\int_0^t \|\alpha_s\|^2 ds < \infty$ a.s. for all $t \geq 0$.

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 $(Z_t)_{t \geq 0}$ is a local martingale, as it is a stochastic integral with respect to a Brownian motion.

Recall that since Z is a positive local martingale, it is automatically a supermartingale. Hence, if

$$\mathbb{E}(Z_T) = 1$$

for some non-random $T > 0$, then $(Z_t)_{0 \leq t \leq T}$ is a true martingale. In this case, what happens to the Brownian motion when we change to an equivalent measure with density Z_T?

THEOREM (Cameron–Martin–Girsanov Theorem). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which a m-dimensional Brownian motion $(W_t)_{t \geq 0}$ is defined, and let $(\mathcal{F}_t)_{t \geq 0}$ 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 $(Z_t)_{0 \leq t \leq T}$ is a martingale. Define the equivalent measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$ by the density process

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

Then the m-dimensional process $(\hat{W}_t)_{0 \leq t \leq T}$ defined by

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

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

REMARK. This isn't the appropriate place to prove Girsanov's theorem, but here is a related result that is completely elementary: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let W be a random vector with the d-dimensional normal $N_d(0, I)$ distribution, where I is the $d \times d$ identity matrix. Fix a constant vector $\alpha \in \mathbb{R}^d$ and define an equivalent measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ by the density

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{\alpha \cdot W - \|\alpha\|^2/2}.$$

Then the random vector $\hat{W} = W - \alpha$ has the $N_d(0, I)$ distribution under $\mathbb{Q}$.

To see why, let $f : \mathbb{R}^d \to \mathbb{R}$ be bounded and measurable. The following computation proves the claim

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}[f(W)] &= \mathbb{E}^{\mathbb{P}}[e^{\alpha \cdot W - \|\alpha\|^2/2} f(W - \alpha)] \\
&= \int e^{\alpha \cdot w - \|\alpha\|^2/2} f(w - \alpha) \frac{e^{-\|w\|^2/2}}{(2\pi)^{d/2}} dw \\
&= \int \frac{e^{-\|w-\alpha\|^2/2}}{(2\pi)^{d/2}} f(w - \alpha) dw \\
&= \int f(u) \frac{e^{-\|u\|^2/2}}{(2\pi)^{d/2}} du \\
&= \mathbb{E}^{\mathbb{P}}[f(W)]
\end{aligned}$$

where the second to last line follows from the change of variables $u = w - \alpha$.

Now, you may be asking yourself: When is the process $(Z_t)_{t\geq 0}$ not just a local martingale, but a true martingale?

THEOREM (Novikov's criterion). *If*

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

then

$$\mathbb{E}\left(e^{-\frac{1}{2}\int_0^T \|\alpha_s\|^2 ds + \int_0^T \alpha_s \cdot dW_s}\right) = 1.$$

6. A martingale representation theorem

In this section we will see that all continuous martingales are essentially stochastic integrals with respect to Brownian motion. This will have applications to our continuous-time financial models in the next chapter.

THEOREM (Itô's Martingale Representation Theorem). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which a m-dimensional Brownian motion $W = (W_t)_{t\geq 0}$ is defined, and let the filtration $(\mathcal{F}_t)_{t\geq 0}$ be the filtration generated by W.*

Let $X = (X_t)_{t\geq 0}$ be a continuous local martingale. Then there exists a unique predictable m-dimensional process $(\alpha_t)_{t>0}$ such that $\int_0^t \|\alpha_s\|^2 ds < \infty$ almost surely for all $t \geq 0$ and

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

Furthermore, if $X_t > 0$ for all $t \geq 0$ then there exists a predictable β such that $\int_0^t \|\beta_s\|^2 ds < \infty$ and

$$X_t = X_0 e^{-\frac{1}{2}\int_0^t \|\beta_s\|^2 ds + \int_0^t \beta_s \cdot dW_s}$$

PROOF OF THE SECOND CLAIM. Assuming that

$$dX_t = \alpha_t \cdot dW_t$$

and the positivity of X, apply Itô's formula to get

$$d \log X_t = \frac{\alpha_t}{X_t} \cdot dW_t - \frac{\|\alpha_t\|^2}{2X_t^2} dt$$

so the conclusion follows with $\beta_t = \alpha_t / X_t$. $\square$

REMARK. This isn't the appropriate place to prove the martingale representation theorem, but here is an easy, related result: Let W be a simple, symmetric random walk, so that $\mathbb{P}(W_t - W_{t-1} = \pm 1 | W_0, \ldots, W_{t-1}) = 1/2$ for all $t \geq 1$. Suppose that the filtration is generated by W, then for every martingale M there exists a previsible process $(\theta_t)_{t\geq 1}$ such that

$$M_t = M_0 + \sum_{s=1}^{t} \theta_s (W_s - W_{s-1}).$$

To see why, note that since M_t is $\mathcal{F}_t$-measurable for each t, there exists a function $f_t : \{0,1\}^t \to \mathbb{R}$ such that

$$M_t = f_t(\zeta_1, \ldots, \zeta_t).$$

where $\zeta_s = W_s - W_{s-1}$. We will make use of the identity

$$f_t(\zeta_1, \ldots, \zeta_t) = \frac{1}{2}(\zeta_t + 1)\ f_t(\zeta_1, \ldots, \zeta_{t-1}, 1) + \tfrac{1}{2}(1 - \zeta_t)\ f_t(\zeta_1, \ldots, \zeta_{t-1}, -1).$$

Indeed, the martingale property implies

$$\begin{aligned} M_{t-1} &= \mathbb{E}(M_t | \mathcal{F}_{t-1}) \\ &= \frac{1}{2}\ f_t(\zeta_1, \ldots, \zeta_{t-1}, 1) + \tfrac{1}{2}\ f_t(\zeta_1, \ldots, \zeta_{t-1}, 0). \end{aligned}$$

so that

$$M_t - M_{t-1} = \frac{1}{2}[f_t(\zeta_1, \ldots, \zeta_{t-1}, 1) - f_t(\zeta_1, \ldots, \zeta_{t-1}, -1)]\zeta_t.$$

The desired representation follows from identifying the $\mathcal{F}_{t-1}$-measurable random variable $\theta_t = \frac{1}{2}[f_t(\zeta_1, \ldots, \zeta_{t-1}, 1) - f_t(\zeta_1, \ldots, \zeta_{t-1}, -1)]$.

If this example has you thinking about the completeness of the binomial tree model, then you might not be surprised to learn that the martingale representation theorem plays a role in studying completeness in continuous time.

We conclude this section with a useful result. It is not directly applicable to finance, but simplifies several arguments.

Theorem (Lévy's Characterisation of Brownian Motion). *Let $(X_t)_{t\geq 0}$ be a continuous m-dimensional local martingale such that*

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

Then $(X_t)_{t\geq 0}$ is a standard m-dimensional Brownian motion.

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

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

By Itô's formula,

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

and so $(M_t)_{t\geq 0}$ 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 $(M_t)_{t\geq 0}$ 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

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

The above equation implies that the increment $X_t - X_s$ has the $N_m(0, (t-s)I)$ distribution and is independent of $\mathcal{F}_s$. $\square$

CHAPTER 3

Continuous-time models

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 n-dimensional stochastic processes $P = (P_t^1, \ldots, P_t^n)_{t\geq 0}$ representing the asset prices. This process will be defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions. Furthermore, we will make the following assumption to make use of the Itô calculus developed in the previous chapter.

Assumption. The stochastic process P is assumed to be is an Itô process adapted to $\mathbb{F}$.

Since continuous-time theory has enough complications, we will make the following simplification:

Assumption. There exists a numéraire asset.

In particular, when we discuss arbitrage theory, there is no need to allow the possibility of intermediate consumption.

As before, the investor's controls consist of the n-dimensional process $H = (H_t^1, \ldots, H_t^n)_{t\geq 0}$ where H_t^i and corresponds to the number of shares of asset i held at time t. We will assume that H is self-financing in the continuous time sense:

Definition. A n-dimensional predictable process H such that H is P-integrable[1]is a *self-financing* (pure-investment) strategy iff

$$d(H_t \cdot P_t) = H_t \cdot dP_t$$

WARNING: THIS DEFINITION IS INCOMPLETE in the sense that it does not give rise to interesting arbitrage theory. The reason for the above warning is spelled out below.

2. Admissible strategies

In order to make sense of the stochastic integral defining the wealth, we need to impose a technical integrability condition which holds automatically for continuous processes.

[1]...this means the stochastic integral $\int_0^t H_s \cdot dP_s$ is well-defined, i.e. if $dP_t = b_t dt + \sigma_t dW_t$ then

$$\int_0^t |H_s \cdot b_s| ds < \infty \text{ and } \int_0^t \|\sigma_s^\top H_s\|^2 ds < \infty \text{ a.s. for all } t \geq 0$$

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. Consider a discrete-time market model with two assets $P = (1, S)$ where S is a simple symmetric random walk:

$$S_t = \xi_1 + \ldots + \xi_t$$

where the random variables $\xi_1, \xi_2, \ldots$ are independent and

$$\mathbb{P}(\xi_t = 1) = \mathbb{P}(\xi_t = -1) = 1/2.$$

Obviously this market has no arbitrage as P is a martingale. Nevertheless, let's explore how to approximate an arbitrage in some sense. Given a predictable process π, let

$$\phi_t = \sum_{s=1}^{t-1} (\pi_{s+1} - \pi_s) S_s$$

Then the pair (ϕ, π) defines a self-financing pure investment strategy with associated wealth process

$$X_t = \sum_{s=1}^{t} \pi_s (S_s - S_{s-1}).$$

In particular, $X_0 = 0$.

A simple strategy that resembles an arbitrage is constructed as follows: first define the stopping time

$$\sigma = \inf\{t \geq 0 : S_t > 0\}.$$

and consider the strategy with

$$\pi_t = \mathbb{1}_{\{t \leq \sigma\}}$$

Note that the associated wealth process is $X_t = S_{t \wedge \sigma}$. Since $\sigma < \infty$ a.s., the conclusion is that if you are willing to wait a while, investing in this strategy will result in an almost sure gain $X_\sigma = 1$. But the amount of time you have to wait is very long: one can show that $\mathbb{E}(\sigma) = +\infty$.

One can improve upon the above idea by taking larger and larger bets, effectively 'speeding up the clock'. Indeed, define the stopping time

$$\tau = \inf\{t \geq 0 : \xi_t = 1\}$$

and consider the strategy

$$\pi_t = 2^{t-1} \mathbb{1}_{\{t \leq \tau\}}.$$

In this case, the associated wealth process is

$$X_t = 1 - 2^t \mathbb{1}_{\{t \leq \tau - 1\}}.$$

This is the classical 'martingale' or doubling strategy. Note that $\mathbb{E}(\tau) = 2$, so an investor following this strategy does not have to wait very long on average to realise the gain $X_\tau = 1$. But although τ is small on average, it is not bounded, and hence this strategy does not qualify as an arbitrage.

Example. A technical problem with continuous time models is that events that will happen eventually can be made to happen in bounded time by speeding up the clock.

Consider the market with prices $P = (1, W)$ where W is a Brownian motion. Given predictable process $(\pi_t)_{t\in[0,T]}$ on non-random horizon $T > 0$, such that

$$\int_0^T \pi_s^2 ds < \infty$$

we can define

$$X_t = \int_0^t \pi_s \, dWs$$

and

$$\phi_t = X_t - \pi_t W_t$$

for $0 \le t \le T$. Note that the strategy $H = (\phi, \pi)$ that $X_t = H_t \cdot P_t$ is the wealth process and satisfies the self-financing condition

$$dX_t = \pi_t dW_t = H_t \cdot dP_t$$

with initial wealth $X_0 = 0$.

We will now construct a process π such that $X_T = K$ a.s. where the constant $K > 0$ is arbitrary.

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

Now define $(Z_u)_{u\ge 0}$ by

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

Note that Z is a local martingale in the filtration $(\mathcal{F}_{f^{-1}(u)})_{u\ge 0}$ and 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 characterisation $(Z_u)_{u\ge 0}$ is a Brownian motion. Define the stopping time τ by

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

Since $(Z_u)_{u\ge 0}$ is a Brownian motion, we have $\tau < \infty$ almost surely since $\sup_{u\ge 0} Z_u = \infty$ almost surely.

Now let

$$\pi_t = (f'(t))^{1/2} \mathbb{1}_{\{t \le f^{-1}(\tau)\}}$$

and note

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

the stochastic integral defining X makes sense. The strange fact is that $(X_t)_{t\in[0,T]}$ is a local martingale with $X_0 = 0$, but $X_T = Z_\tau = K$ almost surely.

We see that integrand $(\pi_s)_{s\in[0,T]}$ roughly 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 integrability necessary to define the stochastic integral is not really sufficient for our needs.

At this stage, there are several reasonable options. **In this course we will insist that the investor's portfolio always has non-negative value.**

Definition. A trading strategy H is *admissible* iff

$$H_t \cdot P_t \geq 0 \text{ for all } t \geq 0 \text{ almost surely }.$$

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 local martingale deflators

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

Definition. An admissible strategy H is called an absolute arbitrage iff there is a non-random time T such that

$$H_0 \cdot P_0 = 0 \leq H_T \cdot P_T \text{ a.s.}$$

and

$$\mathbb{P}\left(H_T \cdot P_T > 0\right) > 0.$$

An admissible strategy H is called an arbitrage relative to a strategy K iff there is a non-random time T such that

$$H_0 \cdot P_0 = K_0 \cdot P_0,$$

and

$$H_T \cdot P_T \geq K_T \cdot P_T \text{ a.s.}, \mathbb{P}\left(H_T \cdot P_T > K_T \cdot P_T\right) > 0.$$

Remark. Note that if H is an absolute arbitrage and K is admissible, then the strategy $H + K$ is an arbitrage relative to K. On the other hand, if H is an arbitrage relative to K, then $H - K$ is an absolute arbitrage only if $H - K$ is admissible. In particular, an absolute arbitrage is an arbitrage relative the strategy $K = 0$ of holding no assets.

In discrete time, the notions of absolute arbitrage and relative arbitrage are essentially equivalent since we did not have to worry about admissibility. In continuous time, we will soon find examples of the surprising fact that

$$\boxed{\text{there exist continuous-time markets that have relative arbitrage but no absolute arbitrage.}}$$

Such market models are sometimes considered models of price bubbles.

The point of all of this is to warn you to be careful when making arbitrage arguments in continuous time, since reasonable people can disagree on what kind of strategies should be called arbitrages.

As in the discrete-time theory, we now introduce martingale deflators.

Definition. A (local) martingale deflator is a positive Itô process Y such that $YP = (Y_t P_t)_{t\geq 0}$ is an n-dimensional (local) martingale.

Our continuous-time version of the first fundamental theorem follows. Unfortunately, to get a clean statement of this result we need to up the technical ante.

Theorem. *Suppose there exists a local martingale deflator for the market model P. If K is a strategy such that the process $K \cdot PY$ is a true martingale, then there is no arbitrage relative to K. In particular, there is no absolute arbitrage.*

The proof of this fact is based on an important lemma:

Lemma. *Suppose H is a self-financing pure investment strategy and let*

$$X_t = H_t \cdot P_t = X_0 + \int_0^t H_s \cdot dP_s$$

Then

$$d(X_t Y_t) = H_t \cdot d(Y_t P_t).$$

for any Itô process Y. In particular, if Y is a local martingale deflator and H is admissible then XY is a supermartingale.

Proof of lemma. . First note

$$Y_t dX_t = Y_t(H_t \cdot dP_t)$$

and

$$X_t dY_t = H_t \cdot P_t dY_t.$$

Finally, note that

$$d\langle X, Y\rangle_t = (H_t \cdot dP_t)(dY_t) = \sum_{i=1}^n H_t^i d\langle P^i, Y\rangle_t.$$

Putting this together with Itô's formula yields

$$\begin{aligned} d(X_t Y_t) &= Y_t dP_t + X_t dY_t + d\langle X, Y\rangle_t \\ &= \sum_i H_t^i (Y_t dP_t^i + P_t^i dY_t + \langle Y, P^i\rangle_t) \\ &= H_t \cdot d(Y_t P_t) \end{aligned}$$

as claimed. Now if Y is a local martingale deflator, then PY is a local martingale. In particular the process XY can be expressed as the stochastic integral with respect to a continuous local martingale, and hence is itself a local martingale. Finally, if H is admissible, then XY is a non-negative local martingale. Non-negative local martingales are supermartingales by Fatou's lemma. ☐

Proof that existence of a local martingale deflator implies no arbitrage. Let Y be a local martingale deflator, and let H and K be strategies such that

$$H_0 \cdot P_0 = K_0 \cdot P_0 \text{ and } H_T \cdot P_T \geq K_T \cdot P_T.$$

Furthermore, suppose that H is admissible and $K \cdot PY$ is a martingale. We must show that $H_T \cdot P_T = K_T \cdot P_T$.

By the above lemma and since Y is non-negative, the process $H \cdot PY$ is a super-martingale. Hence

$$\begin{aligned} K_0 \cdot P_0 Y_0 &= H_0 \cdot P_0 Y_0 \\ &\geq \mathbb{E}(H_T \cdot P_T Y_T) \\ &\geq \mathbb{E}(K_T \cdot P_T Y_T) \\ &= K_0 \cdot P_0 Y_0. \end{aligned}$$

This shows that $H_T \cdot P_T Y_T = K_T \cdot P_T Y_T$. Since Y is strictly positive, the conclusion now follows. □

Remark. Note that the above theorem doesn't say that no relative arbitrage implies the existence of a local martingale deflator. A weaker version notion of relative arbitrage, called 'free-lunch-with-vanishing-risk,' is needed to have the converse implication. See the recent book of Delbaen and Schachermayer *The Mathematics of Arbitrage* for an account of the modern theory.

Remark. Here is an example of a market with a relative arbitrage and no absolute arbitrage. Fix $T > 0$ and let $(\pi_t)_{0 \leq t \leq T}$ be a predictable process such that $\int_0^T \pi_s^2 ds < \infty$ and let $S_t = \int_0^t \pi_s dW_s$ where W is a Brownian motion. Suppose $S_t \leq 1$ for all $t \leq T$ and $S_T = 1$ almost surely. (See the previous section on doubling strategies for an explicit construction of such a process π.)

Now consider the market with prices $P = (1, S)$. Note that P is a two-dimensional local martingale, hence there exists a martingale deflator – just set $Y_t = 1$ for all t. Therefore, there is no absolute arbitrage. However, consider the strategy $K = (1, -1)$. Note that $K_t \cdot P_t = 1 - S_t \geq 0$ so K is admissible. We will show that there exists an arbitrage relative to K. Indeed, let $H = (1, 0)$. Note that $H_0 \cdot P_0 = 1 = K_0 \cdot P_0 = 1$ but $H_T \cdot P_T = 1 > K_T \cdot P_T = 0$.

The point of this example is that the asset with price S seems like a good deal - it costs nothing at time 0 but pays a positive amount at time 1. However, holding one share of the asset, corresponding to the strategy $(0, 1) = H - K$ is not admissible.

4. The structure of local martingale deflators

In this section we will parametrise a fairly general Itô market with $n = d + 1$ assets. All assets in this market are numéraires, and we use the notation $P = (B, S)$. We will assume the dynamics of the prices are given by the following equations

$$\begin{aligned} dB_t &= B_t r_t dt \\ dS_t^i &= S_t^i \left(\mu_t^i dt + \sum_{j=1}^m \sigma_t^{ij} dW_t^j \right) \text{ for } i = 1, \ldots, d \end{aligned}$$

where the processes r, μ^i, σ^{ij} are predictable and suitably integrable, and the W^j are independent Brownian motions.

The first asset can be thought of as a bank account, and the random variable r_t is the spot interest rate at time t. The (random) ordinary differential equation can be solved:

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

The d assets can be thought of as risky stocks. The random variable μ_t^i is interpreted as the mean instantaneous return of asset i, while the spot volatility is $(\sum_j (\sigma_t^{ij})^2)^{1/2}$. Note that Itô's formula yields

$$S_t^i = S_0^i e^{\int_0^t [\mu_s^i - \frac{1}{2}\sum_j (\sigma_s^{ij})^2] ds + \int_0^t \sum_j \sigma_s^{ij} dW_s^j}.$$

We will use the notation

$$\mu_t = \begin{pmatrix} \mu_t^1 \\ \vdots \\ \mu_t^d \end{pmatrix} \text{ and } \sigma_t = \begin{pmatrix} \sigma_t^{11} & \cdots & \sigma_t^{1m} \\ \vdots & \ddots & \vdots \\ \sigma_t^{d1} & \cdots & \sigma_t^{dm} \end{pmatrix}$$

for the $d \times 1$ vector of means and $d \times m$ matrix of volatilities, respectively.

With this more explicit parametrisation, we can describe the structure of local martingale deflators:

Theorem. *Let λ be a predictable m-dimensional process such that $\int_0^t \|\lambda_s\|^2 ds < \infty$ a.s. for all $t \geq 0$ and that*

$$\sigma_t \lambda_t = \mu_t - r_t \mathbf{1} \text{ for almost all } (t, \omega)$$

where $\mathbf{1} = (1, \cdots, 1)^\top$ is the $d \times 1$ vector with the constant 1 in each component.

Let

$$Y_t = Y_0 e^{-\int_0^t (r_s + \|\lambda_s\|^2/2) ds - \int_0^t \lambda_s \cdot dW_s}$$

for a constant $Y_0 > 0$ – or in equivalent differential form

$$dY_t = Y_t(-r_t dt - \lambda_t \cdot dW_t).$$

Then Y is a local martingale deflator.

Furthermore, if the filtration is generated by the m-dimensional Brownian motion W, all local martingale deflators have this form.

Remark. The m-dimensional random vector λ_t appearing the theorem is a generalisation of the Sharpe ratio. The process $\lambda = (\lambda_t)_{t \geq 0}$ is often called the *market price of risk*, for the local martingale deflator, since it measures in some sense the excess return of the stocks per unit of volatility.

Proof. We need to show that YB and YS are local martingales. Note that by Itô's formula

$$d(Y_t B_t) = -Y_t B_t \lambda_t \cdot dW_t$$

so YB is a local martingale since it is the stochastic integral with respect to a Brownian motion W.

Also, by Itô's formula

$$\begin{aligned} d(Y_t S_t^i) &= Y_t S_t^i[-r_t + \mu_t^i - (\sigma_t \lambda_t)^i] + Y_t S_t^i(\sigma_t^{i\cdot} - \lambda_t) \cdot dW_t \\ &= Y_t S_t^i(\sigma_t^{i\cdot} - \lambda_t) \cdot dW_t \end{aligned}$$

where we have used the identity $\sigma_t \lambda_t = \mu_t - r_t \mathbf{1}$ to cancel the dt term.

Conversely, if the filtration is generated by the Brownian motion, the martingale representation theorem says that all positive local martingales M are of the form

$$M_t = M_0 e^{-\frac{1}{2}\int_0^t \|\lambda_s\|^2 ds - \int_0^t \lambda_s \cdot dW_s}$$

for some predictable λ, or in differential form

$$dM_t = -M_t \lambda_t \cdot dW_t.$$

Hence, if $YB = M$ is a positive local martingale then

$$dY_t = -Y_t(r_t dt + \lambda_t \cdot dW_t)$$

by Itô's formula. Furthermore, if YS is a local martingale, then Itô's formula shows that in order to cancel the drift we must have the identity $\sigma_t \lambda_t = \mu_t - r_t \mathbf{1}$. □

5. Replication of European claims

As before, given a market model P we can introduce a contingent claim. Recall that a European contingent claim maturing at a time $T > 0$ is modelled as random variable ξ_T that is $\mathcal{F}_T$-measurable. We shall assume that there exists at least one martingale deflator, so that, in particular, there are no absolute arbitrages.

First a simple result:

PROPOSITION. *Suppose H is an admissible replication strategy for a European contingent claim with time T payout ξ_T, and let Y be a local martingale deflator. Then*

$$H_t \cdot P_t \geq \frac{1}{Y_t} \mathbb{E}(\xi_T Y_T | \mathcal{F}_t).$$

PROOF. This is the same as the proof of that the existence of a local martingale deflator implies no arbitrage.

$$\begin{aligned} \mathbb{E}(\xi_T Y_T | \mathcal{F}_t) &= \mathbb{E}(H_T \cdot P_T Y_T | \mathcal{F}_t) \\ &\leq H_t \cdot P_t Y_t \end{aligned}$$

since $H \cdot PY$ is a supermartingale. □

Now we will impose more structure, by assuming that the market model $P = (B, S)$ has dynamics

$$\begin{aligned} dB_t &= B_t r_t dt \\ dS_t^i &= S_t^i \left(\mu_t^i dt + \sum_{j=1}^m \sigma_t^{ij} dW_t^j \right) \text{ for } i = 1, \ldots, d \end{aligned}$$

as before, or in vector notation, these equations can be written as

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

where

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

We will work in the filtration generated by W, so that all state price densities Y are of the form

$$dY_t = Y_t(-r_t dt - \lambda_t \cdot dW_t).$$

where

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

The following will serve as a version of the second fundamental theorem of asset pricing in continuous time.

THEOREM. *Suppose the filtration is generated by W, and suppose $m = d$ and that the $d \times d$ matrix σ_t is invertible for all (t, ω), so that in particular, there is a unique (up to scaling) martingale deflator Y of the form*

$$dY_t = Y_t(-r_t dt - \lambda_t \cdot dW_t).$$

where

$$\lambda_t = \sigma_t^{-1}(\mu_t - r_t \mathbf{1}).$$

Let ξ_T be non-negative, $\mathcal{F}_T$-measurable and such that $\xi_T Y_T$ is integrable. Then there exists an admissible strategy H such that

$$H_t \cdot P_t = \frac{1}{Y_t} \mathbb{E}(Y_T \xi_T | \mathcal{F}_t).$$

In particular, the strategy H replicates $\xi_T = H_T \cdot P_T$, and the initial cost $\mathbb{E}(Y_T \xi_T)/Y_0 = H_0 \cdot P_0$ is the minimum cost among admissible replication strategies.

REMARK. Clearly, you could also replicate the claim with an admissible strategy by running a suicide strategy on top of the replication strategy H, but the initial cost of this strategy is more. On the other hand, if you didn't care about admissibility, you could employ a doubling strategy to replicate the claim with strictly smaller initial cost.

PROOF. Let

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

Then M is a martingale, and since the filtration is generated by the Brownian motion W the martingale representation theorem tells us that there exists a d-dimensional predictable process α such that

$$dM_t = \alpha_t \cdot dW_t.$$

We will show that there exists an admissible self-financing strategy H such that $X_0 = M_0/Y_0$ and

$$d(X_t Y_t) = \alpha_t \cdot dW_t.$$

where $X_t = H_t \cdot P_t$ is the wealth process. This will show that $X_t = M_t/Y_t$, and hence $X_T = \xi_T$ as claimed.

Now, let H be a self-financing strategy. By an Itô's formula calculation, we have

$$d(X_t Y_t) = H_t \cdot d(Y_t P_t).$$

Write $H = (\phi, \pi)$, where ϕ is the number of shares of the bank account and π^i is the number of shares of stock i. We need the following to hold

$$\phi_t B_t + \pi_t \cdot S_t = M_t/Y_t$$

and

$$\phi_t d(Y_t B_t) + \pi_t \cdot d(Y_t S_t) = \alpha_t \cdot dW_t$$

We can solve for ϕ in terms of π:

$$\phi_t = \frac{1}{B_t}\left(\frac{M_t}{Y_t} - \pi_t \cdot S_t\right).$$

Finally, we know that

$$d(B_t Y_t) = -B_t Y_t \lambda_t \cdot dW_t$$

and

$$d(S_t^i Y_t) = S_t^i Y_t (\sigma_t^{i\cdot} - \lambda_t) \cdot dW_t$$

Plugging this in, yields

$$\pi_t = \mathrm{diag}(S_t)^{-1} (\sigma_t^\top)^{-1} \frac{(M_t \lambda_t + \alpha_t)}{Y_t}.$$

To sum up, note that $\phi_t B_t + \pi_t \cdot S_t = M_t/Y_t$ and that

$$\phi_t dB_t + \pi_t \cdot dS_t = d\left(\frac{M_t}{Y_t}\right).$$

This means $H = (\phi, \pi)$ is a self-financing strategy and

$$H_t \cdot P_t = \frac{M_t}{Y_t} \text{ for all } 0 \le t \le T.$$

It is admissible since $\xi_T \ge 0$ and hence $M_t \ge 0$.

Now, given the existence of the replication strategy H, the minimality of the replication cost follows from the proposition at the beginning of this section since YX is a true martingale (not just a supermartingale). $\square$

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

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

6. Risk-neutral measures

As before, we have $n = 1 + d$ asset withe prices $P = (B, S)$ where

$$
\begin{aligned}
dB_t &= B_t r_t dt \\
dS_t &= \mathrm{diag}(S_t)(\mu_t dt + \sigma_t dW_t).
\end{aligned}
$$

In this section we assume that there exists a local martingale deflator Y.

DEFINITION. A risk-neutral measure is an equivalent probability measure $\mathbb{Q}$ such that S/B is a $\mathbb{Q}$-local martingale.

Note that YB is a local martingale by the definition of a martingale deflator. Assuming that YB is a true martingale, then we can construct a risk-neutral measure $\mathbb{Q}$ as we have done previously by fixing $T > 0$ and setting

$$
\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{B_T Y_T}{B_0 Y_0}.
$$

Conversely, if $\mathbb{Q}$ is a risk-neutral measure, we have can construct a local martingale deflator Y by setting

$$
Y_t = \frac{1}{B_t} \mathbb{E}\left(\frac{d\mathbb{Q}}{d\mathbb{P}} | \mathcal{F}_t \right).
$$

Note that by this construction YB is a true martingale. This leads to the following result. It is just an application of our continuous time version of the 1FTAP:

PROPOSITION. *Suppose there exists a risk-neutral measure, then there is no arbitrage relative to the bank account.*

Now let's explore what the asset prices look like under a risk-neutral measure.

PROPOSITION. *Let $\mathbb{Q}$ be a risk-neutral measure, and suppose that the filtration is generated by the Brownian motion. Then the asset prices have dynamics*

$$
\begin{aligned}
dB_t &= B_t r_t dt \\
dS_t &= \mathrm{diag}(S_t)(r_t \mathbf{1} dt + \sigma_t d\hat{W}_t).
\end{aligned}
$$

where $\hat{W}$ is a Brownian motion under $\mathbb{Q}$.

PROOF. By the structure theorem for local martingale deflators we have that any martingale deflator Y satisfies

$$
\frac{Y_t B_t}{Y_0 B_0} = e^{-\int_0^t \|\lambda_s\|^2 ds - \int_0^t \lambda_s \cdot dW_s}.
$$

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

Now if there exists a risk-neutral measure $\mathbb{Q}$, then YB is a true martingale. Hence we may apply Girsanov's theorem to conclude that the process $\hat{W}$ defined by

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

is a Brownian motion under $\mathbb{Q}$. Note

$$\begin{aligned} dS_t &= \mathrm{diag}(S_t)(\mu_t dt + \sigma_t dW_t) \\ &= \mathrm{diag}(S_t)[\mu_t dt + \sigma_t(d\hat{W}_t - \lambda_t dt)] \\ &= \mathrm{diag}(S_t)(r_t \mathbf{1} dt + \sigma_t d\hat{W}_t) \end{aligned}$$

as claimed. □

We finally connect risk-neutral measures to replication costs. This is just restatement of the version of the 2FTAP appearing the previous section:

Proposition. *Assume that there exists a risk-neutral measure $\mathbb{Q}$, that the filtration is generated by W, and suppose $m = d$ and that the $d \times d$ matrix σ_t is invertible for all (t, ω).*

Let ξ_T be non-negative, $\mathcal{F}_T$-measurable and such that ξ_T / B_T is $\mathbb{Q}$-integrable. Then there exists an admissible strategy H that replicates the claim $\xi_T = H_T \cdot P_T$, and the initial cost $\mathbb{E}^{\mathbb{Q}}(e^{-\int_0^T r_s ds} \xi_T) = H_0 \cdot P_0$ is minimum among admissible replication strategies.

7. The Black–Scholes model and formula

We will consider the simplest possible model of the type studied introduced above. 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 $\sigma > 0$. We will assume that the filtration is generate by the scalar Brownian motion W. This is often called the *Black–Scholes* model.

We are interested in finding the replication cost of a European contingent claim with payout $\xi_T = g(S_T)$, where g is a given function which we assume to be non-negative and suitably integrable. We know from before that the unique state price density with $Y_0 = 1$ is given by

$$Y_t = e^{-(r+\lambda^2/2)t - \lambda W_t}$$

where $\lambda = (\mu - r)/\sigma$.

Hence, from our existential result there is a trading strategy H which replicates the payout with time t cost

$$X_t = \frac{1}{Y_t} \mathbb{E}[Y_T g(S_T) | \mathcal{F}_t].$$

This is where we see the advantage of working with equivalent martingale measures rather than state price densities. Indeed, define the equivalent martingale measure $\mathbb{Q}$ by the density

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{-\lambda^2 T/2 - \lambda W_T}$$

and recall that by the Cameron–Martin–Girsanov theorem the process $\hat{W}_t = W_t + \lambda t$ is a $\mathbb{Q}$-Brownian motion.

The price of the stock can be written explicitly:

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

and hence

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

A famous example is the case of the European call option where the payout function is of the form $g(S) = (S-K)^+$. In this case, we have the the Nobel-prize-winning *Black–Scholes formula*:

$$
\begin{aligned}
C_t(T,K) =& S_t\Phi\left(-\frac{\log(K/S_t)}{\sigma\sqrt{T-t}} + (r/\sigma+\sigma/2)\sqrt{T-t}\right) \\
&- Ke^{-r(T-t)}\Phi\left(-\frac{\log(K/S_t)}{\sigma\sqrt{T-t}} + (r/\sigma-\sigma/2)\sqrt{T-t}\right)
\end{aligned}
$$

where $\Phi(x) = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy$ is the standard normal distribution function. (You are asked to derive this formula on Example Sheet 3.)

We have argued that the martingale representation theorem asserts the existence of replicating strategy H, but unfortunately, it gives us no information about how to compute H. This problem will be tackled in the next section.

8. Markovian markets: pricing and hedging by PDE

We now have a sufficient condition that a contingent claim can be replicated. 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 this section.

The first step is to pose a model for the asset prices $(B_t, S_t)_{t\geq 0}$. A good model should give a reasonable statistical fit to the actual market data. 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

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

where the nonrandom functions $r : [0,\infty)\times\mathbb{R}^d \to \mathbb{R}$, $\mu : [0,\infty)\times\mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0,\infty)\times\mathbb{R}^d \to \mathbb{R}^{d\times m}$ are given. Notice that this is a special case of the set-up of the last section, as now (with an abuse notation)

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

In this special situation, the asset prices $(S_t)_{t\geq 0}$ are a d-dimensional Markov process.

The next theorem says how to find a replicating strategy for a contingent claim maturing at time T with payout

$$\xi_T = g(S_T)$$

for some non-random function $g : \mathbb{R}^d \to [0, \infty)$.

THEOREM. *Suppose the function $V : [0,T] \times \mathbb{R}^d \to [0, \infty)$ satisfies the partial differential equation*

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

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

Then there exists an-admissible strategy H such that $H_t \cdot P_t = V(t, S_t)$. In particular, this strategy replicates the contingent claim with payout $g(S_T)$.

Furthermore, if $H = (\phi, \pi)$ then the strategy can be calculated as

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

and

$$\phi_t = \frac{V(t, S_t) - \pi_t \cdot S_t}{B_t}.$$

The above theorem says that if the market model is Markovian, the price (i.e. replication cost) 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 gradient of the pricing function V with respect to the spatial variables, evaluated at time t and current price S_t.

PROOF. By Itô's formula we have

$$\begin{aligned}
dV(t, S_t) &= \frac{\partial V}{\partial t} dt + \sum_i \frac{\partial V}{\partial S^i} dS_t^i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 V}{\partial S^i \partial S^j} d\langle S^i, S^j \rangle_t \\
&= \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sum_{i,j} \frac{\partial^2 V}{\partial S^i \partial S^j} S^i S^j a^{ij} \right) dt + \sum_i \frac{\partial V}{\partial S^i} dS_t^i \\
&= r \left(V - \sum_i S^i \frac{\partial V}{\partial S^i} \right) dt + \sum_i \frac{\partial V}{\partial S^i} dS_t^i
\end{aligned}$$

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

where we have used the assumption that V solves a certain PDE to go from the second to third line above.

Now letting ϕ and π be as in the statement of the theorem we have that

$$
\begin{aligned}
V(t, S_t) &= \phi_t B_t + \pi_t \cdot S_t \\
dV(t, S_t) &= \phi_t dB_t + \pi_t \cdot dS_t.
\end{aligned}
$$

Hence $H = (\phi, \pi)$ is a self-financing strategy with associated wealth process $X_t = V(t, S_t)$ as claimed. It is admissible since $V \geq 0$ by assumption. $\square$

9. Partial differential equations and conditional expectations

We have seen that there are two distinct ways to find a replication cost of a attainable contingent claim: by computing expectations or by solving a PDE. On the one hand, we have shown that the minimal replication cost is the expected value of the discounted payout under a risk-neutral measure. On the other hand, we have seen that if the payout is of the form $g(S_T)$ and the prices $(S_t)_{0 \leq t \leq T}$ satisfy a stochastic differential equation (SDE), then a (not necessarily minimal) replication cost can be found by solving a partial differential equation, and indeed, the replicating strategy is given by the gradient of the solution of the PDE.

In this section we consider the connection between stochastic differential equations and partial differential equations.

The main idea for this section is contained in this result:

THEOREM. *Let the n-dimensional process Z satisfies the SDE*

$$dZ_t = b(Z_t)dt + \sigma(Z_t)dW_t$$

where W is a m-dimensional Brownian motion. Given function f and g, suppose $v : [0, T] \times \mathbb{R}^n \to \mathbb{R}$ is C^2, bounded and satisfies the PDE

$$\frac{\partial v}{\partial t} + \sum_i b^i \frac{\partial v}{\partial z^i} + \frac{1}{2} \sum_{i,j} a^{ij} \frac{\partial^2 v}{\partial z^i \partial z^j} = fV$$

where

$$a^{ij} = \sum_k \sigma^{ik} \sigma^{jk}$$

with terminal condition

$$v(T, z) = g(z) \text{ for all } z \in \mathbb{R}^n.$$

Let

$$M_t = e^{-\int_0^t f(Z_s)ds} v(t, Z_t) \text{ for } 0 \leq t \leq T.$$

Then M is a local martingale. If M is a true martingale (for instance, if M is bounded) then

$$v(t, z) = \mathbb{E}\left[e^{-\int_t^T f(X_s)ds} g(Z_T) | Z_t = z\right].$$

PROOF. By Itô's formula and the fact that v satisfies a certain PDE, we have

$$dM_t = e^{-\int_0^t f(Z_s)ds} \sum_{ij} \sigma^{ij} \frac{\partial v}{\partial z^i} dW^j.$$

Hence M is a local martingale.

If M is true martingale, we have

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

and hence

$$v(t, Z_t) = e^{\int_0^t f(Z_s)ds}\mathbb{E}\left[e^{-\int_0^T f(Z_s)ds}g(Z_T)|\mathcal{F}_t\right].$$

By applying slot property to move the exponential inside the expectation on the right-hand side, then computing the conditional expectation of both sides Z_t and applying the tower and finishes the argument. (Technical note: we have implicitly assumed that all random variables appearing in the conditional expectations are integrable.) □

There is a certain converse to this result.

Theorem. *Let the n-dimensional Markov process Z is a solution of the SDE*

$$dZ_t = b(Z_t)dt + \sigma(Z_t)dW_t$$

where b and σ are continuous and where W is a m-dimensional Brownian motion. Further, assume

$$\mathbb{P}(Z_t \in A|Z_0 = z) > 0$$

for any open set $A \subseteq \mathbb{R}^n$ and any starting point $z \in \mathbb{R}^n$ and any time $t > 0$. Fix a non-random time horizon $T > 0$ and functions f and g and let

$$v(t, z) = \mathbb{E}\left[e^{-\int_t^T f(Z_s)ds}g(Z_T)|Z_t = z\right],$$

assuming sufficient integrability that the conditional expectation is well-defined. If the function is v is twice-continuously differentiable, then v satisfies the PDE

$$\frac{\partial v}{\partial t} + \sum_i b^i \frac{\partial v}{\partial z^i} + \frac{1}{2}\sum_{i,j} a^{ij}\frac{\partial^2 v}{\partial z^i \partial z^j} = fV$$

where

$$a^{ij} = \sum_k \sigma^{ik}\sigma^{jk}$$

with terminal condition

$$v(T, z) = g(z) \text{ for all } z \in \mathbb{R}^n.$$

Proof. By the Markov property

$$\begin{aligned} v(t, Z_t) &= \mathbb{E}\left[e^{-\int_t^T f(Z_s)ds}g(Z_T)|Z_t\right] \\ &= \mathbb{E}\left[e^{-\int_t^T f(Z_s)ds}g(Z_T)|\mathcal{F}_t\right] \end{aligned}$$

where $(\mathcal{F}_t)_{t\geq 0}$ is the filtration generated by Z. Hence, the process M defined by

$$\begin{aligned} M_t &= e^{-\int_0^t f(Z_s)ds}v(t, Z_t) \\ &= \mathbb{E}\left[e^{-\int_0^T f(Z_s)ds}g(X_T)|\mathcal{F}_t\right] \end{aligned}$$

is a martingale. By assumption, the function v is twice-continuously differentiable, so that Itô's formula is applicable:

$$\begin{aligned} dM_t =& e^{-\int_0^t f(X_s)ds} \sum_{ij} \sigma^{ij} \frac{\partial v}{\partial z^i} dW^j \\ &+ e^{-\int_0^t f(Z_s)ds} \left(\frac{\partial v}{\partial t} + \sum_i b^i \frac{\partial v}{\partial z^i} + \frac{1}{2} \sum_{i,j} a^{ij} \frac{\partial^2 v}{\partial z^i \partial z^j} - fV \right) dt \end{aligned}$$

Since M is a martingale, the drift must vanish for almost every (t, ω). And since v is C^2 and Z can hit every open set arbitrarily quickly, we conclude that the drift vanishes identically. □

Remark. The partial differential equation

$$\frac{\partial v}{\partial t} + \sum_i b^i \frac{\partial v}{\partial z^i} + \frac{1}{2} \sum_{i,j} a^{ij} \frac{\partial^2 v}{\partial z^i \partial z^j} = fV$$

is called the Feynman–Kac PDE , whereas the equation

$$v(t, z) = \mathbb{E} \left[e^{-\int_t^T f(Z_s)ds} g(Z_T) | Z_t = z \right]$$

is called the Feynman–Kac formula . The above theorems say, roughly, that a function satisfies the Feynmann–Kac PDE if and only if it satisfies the Feynman–Kac formula.

Example (Black–Scholes continued). Let's return to the Black–Scholes model

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

with constant coefficients r, σ, μ, with $\sigma > 0$. If we would like to replicate a claim with payout $g(S_T)$, the we know that we should solve the so-called *Black–Scholes* PDE

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

But how can we solve this PDE? By the Feynmann–Kac formula!

$$\begin{aligned} V(t, S) &= \mathbb{E}^{\mathbb{Q}}[e^{-r(T-t)} g(S_T) | S_t = S] \\ &= e^{-r(T-t)} \int_{-\infty}^{\infty} g\left(Se^{(r-\sigma^2/2)(T-t)+\sigma\sqrt{T-t}z} \right) \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz \end{aligned}$$

where $\mathbb{Q}$ is the unique risk-neutral measure under which the process $\hat{W}_t = W_t + \lambda t$ is a Brownian motion, where $\lambda = (\mu - r)/\lambda$. Note that

$$dS_t = S_t r \ dt + S_t \sigma \ d\hat{W}_t$$

so the Feynmann–Kac PDE for the Markov process S (when viewed on the risk-neutral measure) is exactly the Black–Scholes PDE.

Now, let's specialise to the case of the call option where $g(S) = (S-K)^+$. From last section we have

$$\begin{aligned} V(t,S) =& S\Phi\left(-\frac{\log(K/S)}{\sigma\sqrt{T-t}} + (r/\sigma+\sigma/2)\sqrt{T-t}\right) \\ &- Ke^{-r(T-t)}\Phi\left(-\frac{\log(K/S)}{\sigma\sqrt{T-t}} + (r/\sigma-\sigma/2)\sqrt{T-t}\right). \end{aligned}$$

The delta, i.e. the replicating portfolio, in this case is (by a miracle of algebra)

$$\frac{\partial V}{\partial S}(t,S) = \Phi\left(-\frac{\log(K/S)}{\sigma\sqrt{T-t}} + (r/\sigma+\sigma/2)\sqrt{T-t}\right).$$

Note that an agent attempting to replicate a call option using the Black–Scholes theory will always hold a fraction of shares of the underlying stock between 0 and 1. Also note that since the sensitivity of the portfolio to the price of the underlying, is given by the formula

$$\frac{\partial^2 V}{\partial S^2}(t,S) = \frac{1}{S\sigma\sqrt{T-t}}\phi\left(-\frac{\log(K/S)}{\sigma\sqrt{T-t}} + (r/\sigma+\sigma/2)\sqrt{T-t}\right)$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. Since the gamma is always positive, the hedger will buy more shares of the underlying if the price goes up.

10. Risk-neutral pricing and PDEs

In this section, we will build an arbitrage-free market model. By the fundamental theorem of asset pricing, a sufficient condition for no-arbitrage is the existence of a risk-neutral measure. Rather than modelling everything under a real-world measure $\mathbb{P}$ and then ensuring that there exists a risk-neutral measure $\mathbb{Q}$, we will simply assume that the measure $\mathbb{Q}$ exists and do all our modelling the on the probability space $(\Omega, \mathcal{F}, \mathbb{Q})$.

We specialise to a certain class of models called factor models. We assume there is a N-dimensional economic factor process Z whose dynamics satisfy a stochastic differential equation

$$dZ_t = b(Z_t)dt + \sigma(Z_t)dW_t$$

where b takes values in $\mathbb{R}^N$ and σ takes values in the $N \times m$ matrices, where W is a Brownian motion in the risk-neutral measure $\mathbb{Q}$. (In previous section we had used the notation W for a Brownian motion under the real-world measure $\mathbb{P}$ and $\hat{W}$ for a Brownian motion under the risk-neutral measure $\mathbb{Q}$. However, in this section, we do not model anything under $\mathbb{P}$, hence we simplify things by dropping the $\hat{}$ from the notation.)

We assume that the spot interest rate $r_t = r(Z_t)$ is a deterministic function of the economic factor. The bank account evolves as

$$dB_t = B_t r_t dt$$

as usual. We build the prices of a second as follows. We consider a function V that iss twice-continuously differentiablea and satisfies the PDE

$$\frac{\partial V}{\partial t} + \sum_i b^i \frac{\partial V}{\partial z^i} + \frac{1}{2}\sum_{i,j} a^{ij}\frac{\partial^2 V}{\partial z^i \partial z^j} = rV$$

where

$$a^{ij} = \sum_k \sigma^{ik}\sigma^{jk}.$$

We then set $P_t = V(t, Z_t)$. By Itô's formula, the discounted price P_t/B_t is a local martingale, and hence there is no arbitrage relative to the bank account.

We now give some examples of risk-neutral pricing in the context of interest rate and stochastic volatility.

In the following interest rate models, we will set $N = 1$ and assume that the economic factor Z_t is the spot interest rate r_t. Our goal will be to price a zero-coupon bond.

10.1. Vasicek model. In 1977, Vasicek proposed the following model for the short rate:

$$dr_t = \lambda(\bar{r} - r_t)dt + \gamma dW_t$$

for a parameter $\bar{r} > 0$ interpreted as a mean short rate, a mean-reversion parameter $\lambda > 0$, and a volatility parameter $\gamma > 0$. This stochastic differential equation can be solved explicitly to yield

$$r_t = e^{-\lambda t}r_0 + (1 - e^{-\lambda t})\bar{r} + \int_0^t e^{-\lambda(t-s)}\gamma dW_s.$$

Note that the short interest rate in the Vasicek model follows an *Ornstein–Uhlenbeck* process, and in particular, that for each $t \geq 0$ the random variable r_t is Gaussian under the measure $\mathbb{Q}$ with

$$\mathbb{E}^{\mathbb{Q}}(r_t) = e^{-\lambda t}r_0 + (1 - e^{-\lambda t})\bar{r} \quad \text{and} \quad \mathrm{Var}^{\mathbb{Q}}(r_t) = \int_0^t e^{-2\lambda(t-s)}\sigma^2 ds = \frac{\gamma^2}{2\lambda}(1 - e^{-2\lambda t}).$$

Moreover, one can show that the process is ergodic and converges to the invariant distribution $N\left(\bar{r}, \frac{\gamma^2}{2\lambda}\right)$. 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 $\mathbb{Q}$-probability of the event $\{r_t < 0\}$ is pretty small.

We have learned from example sheet 3 that

$$\begin{aligned}
\int_0^T r_t dt &= \int_0^T [e^{-\lambda t}r_0 + (1 - e^{-\lambda t})\bar{r}]dt + \int_0^T \int_0^t e^{-\lambda(t-s)}\gamma dW_s \; dt \\
&= \int_0^T [e^{-\lambda t}r_0 + (1 - e^{-\lambda t})\bar{r}]dt + \int_0^T \left[\int_s^T e^{-\lambda(t-s)}dt\right]\gamma dW_s \\
&\sim N\left(\int_0^T [e^{-\lambda t}r_0 + (1 - e^{-\lambda t})\bar{r}]dt, \frac{\gamma^2}{\lambda^2}\int_0^T (1 - e^{-\lambda t})^2 dt\right)
\end{aligned}$$

under $\mathbb{Q}$, so that, using the moment generating function of a Gaussian random variable we have that an initial zero-coupon price can be calculated via risk-neutral expectation:

$$\begin{aligned} P_0^T &= \mathbb{E}^{\mathbb{Q}}[e^{-\int_0^T r_t dt}] \\ &= \exp\left(-\int_0^T \left[e^{-\lambda t} r_0 + (1 - e^{-\lambda t})\bar{r} - \frac{\gamma^2}{2\lambda^2}(1 - e^{-\lambda t})^2\right] dt\right). \end{aligned}$$

The initial forward rates are given by

$$f_0^T = -\frac{\partial}{\partial T} \log P_0^T = e^{-\lambda t} r_0 + (1 - e^{-\lambda t})\bar{r} - \frac{\gamma^2}{2\lambda^2}(1 - e^{-\lambda t})^2$$

By the time-homogeneity of the Vasicek model, we can actually deduce the formula

$$\boxed{f_t^{t+x} = r_t e^{-\lambda x} + \bar{r}(1 - e^{-\lambda x}) - \frac{\gamma^2}{2\lambda^2}(1 - e^{-\lambda x})^2}$$

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.)

We can also compute bond prices by solving a PDE. We save this calculation for the next model:

10.2. Cox–Ingersoll-Ross model. In 1985, Cox, Ingersoll, and Ross proposed the following model for the short rate:

$$dr_t = \lambda(\bar{r} - r_t) + \gamma\sqrt{r_t} dW_t$$

for a parameter $\bar{r} > 0$ interpreted as a mean short rate, a mean-reversion parameter $\lambda > 0$, and a volatility parameter $\gamma > 0$. The process $(r_t)_{t \geq 0}$ 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. This process was also used by Heston to model the spot volatility process in an equity market.

Although the CIR 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 distribution 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.

We can also use the above theorem to compute bond prices. Indeed, fix $T > 0$ and consider the PDE

$$\begin{aligned} &\frac{\partial V}{\partial t}(t, r) + \lambda(\bar{r} - r)\frac{\partial V}{\partial r}(t, r) + \frac{1}{2}\gamma^2 r \frac{\partial^2 V}{\partial r^2}(t, r) = rV(t, r) \\ &V(T, r) = 1. \end{aligned}$$

We can make the log-affine ansatz

$$V(t, r) = e^{rR(T-t) + Q(T-t)}$$

for some functions R and Q which satisfy the boundary conditions $R(0) = Q(0) = 0$. Substituting this into the PDE yields

$$(-\dot{R}r - \dot{Q}) + \lambda(\bar{r} - r)R + \frac{\sigma^2}{2} rR^2 = r$$

Matching coefficients of r yields

$$\begin{aligned}\dot{R} &= -\lambda R + \frac{\sigma^2}{2} R^2 - 1 \\ \dot{Q} &= \lambda \bar{r} R.\end{aligned}$$

The equation for R is a Riccati equation, whose solution is

$$R(\tau) = -\frac{2(e^{\gamma\tau} - 1)}{(\gamma + \lambda)e^{\gamma\tau} + (\gamma - \lambda)}$$

$$Q(\tau) = \int_0^\tau \lambda \bar{r} R(s) ds$$

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.

11. Beyond Black–Scholes

Recall that the Black–Scholes model predicts that the initial price (or more properly, the minimum initial cost of replicating) a European call option of maturity T and strike K is given by the formula

$$C_0(T, K) = S_0 F(\sigma^2 T, K e^{-rT}/S_0)$$

where F is an explicit function (calculated in example sheet 3). What made the Black–Scholes formula so popular after its publication in 1973 is the fact that the right-hand-side depends only on five quantities: the option's maturity time T, the option's strike K, the interest rate r (assumed constant), the underlying stock's initial price S_0 at time t, and a volatility parameter σ. Of these five numbers, only the volatility parameter is neither specified by the option contract nor quoted in the market.

In reality we do not know σ; however we can observe the call prices. So one approach to find the volatility parameter is to observe the prices of calls from the market, and then try to work out which σ to put into the Black–Scholes formula to get the right price.

This is called the *implied volatility* of the option: the unique number $\Sigma(T, K) = \sigma$ such that

$$C_0^{\text{obs}}(T, K) = S_0 F(\sigma^2 T, K e^{-rT}/S_0)$$

If the market was still pricing call options by Black–Scholes formula, then there would exist one parameter σ such that $\Sigma(T, K) = \sigma$ for all T and K. However, in real-world markets, is is usually the case that the implied volatility surface $(T, K) \mapsto \Sigma(T, K)$ is not flat.

Indeed, for fixed T, the graph of the function $K \mapsto \Sigma(T, K)$ often resembles a convex parabola[3] at least for strikes K close to the money, i.e. such that $K e^{-rT}/S_0 \approx 1$. That is

[3]...but be careful: for large K, the graph can grow no faster than $\sqrt{2 \log K/T}$ by a result of Roger Lee.

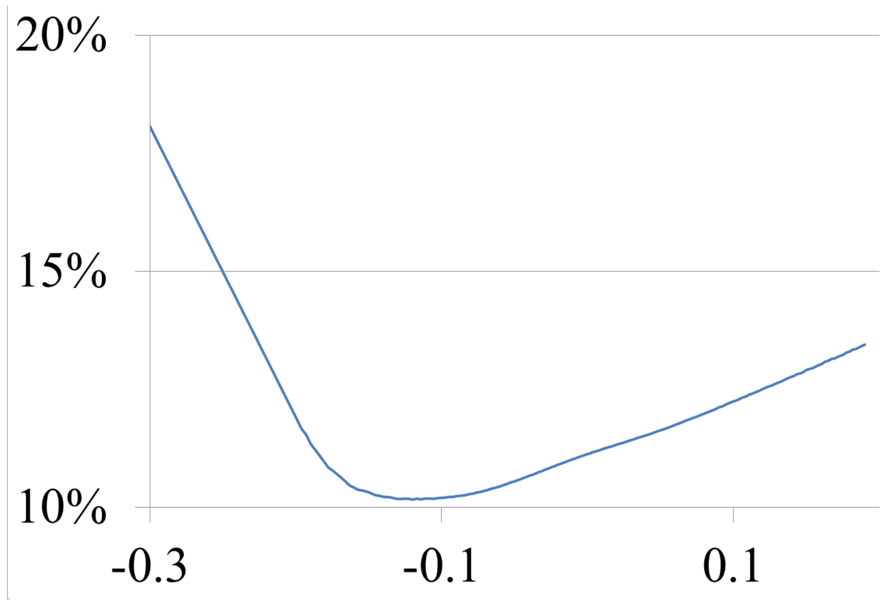


FIGURE 1. Implied volatility vs. log-moneyness $\log(Ke^{-rT}/S_0)$ for E-mini S& P Mid-Cap 400 call options for maturity $T = 0.7$ years, downloaded from `ftp://ftp.cmegroup.com/pub/settle/stleqt` on 12 July 2018

why practitioners refer to the function $K \mapsto \Sigma(T, K)$ as the implied volatility *smile* or *smirk*.

One could either conclude Black–Scholes model is the true model of the stock price and that 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.

11.1. Heston model. In order to explain the observed implied volatility smile, we will then consider a stochastic volatility model. For this model, we will concentrate on the case where there is $d = 1$ stock. For simplicity, we will set the interest rate r to be a constant. It was introduced by Heston in 1993:

$$\begin{aligned} dB_t &= B_t r dt \\ dS_t &= S_t(r dt + \sqrt{v_t} dW_t^S) \\ dv_t &= \lambda(\bar{v} - v_t)dt + c\sqrt{v_t} dW_t^v \end{aligned}$$

Here W^S and W^v are assumed to be correlated Brownian motions *in a fixed risk-neutral measure* $\mathbb{Q}$, with correlation ρ. Correlated Brownian motions can be constructed, for instance, by letting W^v and $W^\perp$ be independent Brownian motions and let

$$W_t^S = \rho W_t^v + \sqrt{1-\rho^2} W_t^\perp.$$

In this model the squared volatility v is a *mean-reverting* process , i.e. an ergodic Markov process, at least under $\mathbb{Q}$. The interpretation of $\bar{v}$ is the level of mean reversion, while λ is the speed of mean reversion. Note that the interest rate is constant r so that the risk-neutral measure $\mathbb{Q}$ is also a forward measure.

Note that Heston's model is special case of the general factor model we considered before, where the factor is $Z_t = (S_t, v_t)$ has dynamic parameters

$$b(S,v) = (rS, \lambda(\bar{v} - v))$$

and

$$\sigma(S,v) = \begin{pmatrix} S\sqrt{v}\rho & S\sqrt{v}\sqrt{1-\rho^2} \\ c\sqrt{v} & 0 \end{pmatrix}.$$

The Heston PDE is then

$$\frac{\partial F}{\partial t} + rS\frac{\partial F}{\partial S} + \lambda(\bar{v} - v)\frac{\partial F}{\partial v} + \frac{1}{2}S^2 v\frac{\partial^2 F}{\partial S^2} + Sv\gamma\rho\frac{\partial^2 F}{\partial S \partial v} + \frac{1}{2}\gamma^2 v\frac{\partial^2 F}{\partial v^2} = rF.$$

Note that $F(t,S,v) = S$ is a solution of the above PDE, so that the market with the bank account and stock has no arbitrage.

We would like to compute call prices $\mathbb{E}[e^{-rT}(S_T - K)^+]$ so we could try to solve Heston's PDE with terminal condition

$$F(T,S,v) = (S-K)^+.$$

While this PDE with boundary conditions can be solved numerically, it seems that no explicit solution is possible, unfortunately.

However, recall that call prices can be calculated via a Fourier integral as long as the moment generating function of the log stock price is known under the forward measure. It turns out that in Heston's model, we can compute the moment generating function reasonably explicitly. To do so, we need to solve Heston's PDE with boundary condition

$$F(T,v,S;\theta) = S^\theta$$

for $\theta \in \Theta = \{p + \mathrm{i}q : 0 < p < 1, q \in \mathbb{R}\} \subseteq \mathbb{C}$. It turns out that this PDE can be solved explicitly. The trick is to make the ansatz

$$F(t,v,S;\theta) = S^\theta e^{R(T-t;\theta)v + Q(T-t;\theta)}.$$

Note that the boundary condition force $R(0;\theta) = Q(0;\theta) = 0$. The PDE becomes

$$-\dot{R}v - \dot{Q} + [\theta r + (\theta^2 - \theta)v/2] + [\lambda\bar{v} + (\theta c\rho - \lambda)v]R + \frac{1}{2}c^2 vR^2 = r,$$

where the dot indicates differentiation with respect to the time variable. Notice that the equation can be written in the form

$$\alpha(T-t;\theta)v + \beta(T-t;\theta) = 0.$$

Now, the above equation should hold for all v so $\alpha(T-t;\theta) = 0 = \beta(T-t;\theta)$, i.e

$$\begin{aligned}\dot{R} &= (\theta^2 - \theta)/2 + (\theta c\rho - \lambda)R + \frac{1}{2}c^2 R^2 \\ \dot{Q} &= (\theta - 1)r + \lambda\bar{v}R.\end{aligned}$$

The equation for R is a Riccati equation which can be solved explicitly. In fact, we do not even have to make any tricky substitutions, separation of variables and partial fractions work well enough:

$$\begin{aligned}
&\dot{R} = \frac{1}{2}c^2(R - R_+)(R - R_-) \\
&\Rightarrow \frac{\dot{R}}{(R - R_+)(R - R_-)} = \frac{1}{2}c^2 \\
&\Rightarrow \frac{1}{R_+ - R_-}\left(\frac{1}{R - R_+} - \frac{1}{R - R_-}\right)\dot{R} = \frac{1}{2}c^2 \\
&\Rightarrow \log\left(\frac{1 - R(\tau)/R_+}{1 - R(\tau)/R_-}\right) = \gamma\tau \\
&\Rightarrow R(\tau;\theta) = (\theta^2 - \theta)\frac{e^{\gamma(\theta)\tau} - 1}{(\gamma(\theta) - \theta c\rho + \lambda)e^{\gamma(\theta)\tau} + (\gamma(\theta) + \theta c\rho - \lambda)}
\end{aligned}$$

where $\gamma(\theta) = \sqrt{(\lambda - \theta c\rho)^2 - (\theta^2 - \theta)c^2}$ and $R_\pm(\theta) = [(\lambda - \theta c\rho)^2 \pm \gamma(\theta)]/c^2$. And the second equation can be solved

$$\begin{aligned}
Q(\tau;\theta) &= \theta r\tau + \int_0^\tau \lambda\bar{v}R(s;\theta)ds \\
&= \left(\theta r + \frac{(\theta^2 - \theta)\lambda\bar{v}}{\gamma(\theta) + \theta c\rho - \lambda}\right)\tau - \frac{2\lambda\bar{v}}{c^2}\log\left(\frac{(\gamma(\theta) - \theta c\rho + \lambda)e^{\gamma(\theta)\tau} + (\gamma(\theta) + \theta c\rho - \lambda)}{2\gamma(\theta)}\right)
\end{aligned}$$

It can be shown that for $\theta \in \Theta$ that

$$\mathbb{E}^{\mathbb{Q}}(e^{\theta\log S_T}) = e^{rT + \theta\log S_0 + R(T;\theta)v_0 + Q(T;\theta)}.$$

What is the point of this calculation? Although the formula for the moment generating function is hard to call beautiful, it is very explicit. In particular, given the set of model parameters $(v_0, \lambda, \bar{v}, c)$, the function can be evaluated very quickly on a computer, and hence the Bromwich integral for call prices can be computed numerically quickly. Hence, it is possible to calibrate the Heston model to market data in a reasonable amount of time. This is one of the main reasons for its popularity.

11.2. Local volatility models. In the previous section we have considered a stochastic volatility model that can match observed call prices reasonably well. However, in this section we will see that it is possible to find a model that matches all observed call prices exactly:

We consider a model given by

$$\begin{aligned}
dB_t &= B_t\, r\, dt \\
dS_t &= S_t(r\, dt + \sigma(t, S_t)d\hat{W}_t).
\end{aligned}$$

That is, the idea is replace the constant volatility parameter in Black–Scholes model with a *local volatility* function $\sigma : [0,\infty) \times (0,\infty) \to (0,\infty)$. We will assume that σ is smooth and bounded from below and above.

The next theorem in the present context is usually attributed to Dupire's 1994 paper.

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. We have already seen a PDE for the replication cost of options in Markovian models. In that PDE, the solution $V(t, S_t)$ was the time-t value of a replication strategy for the given claim, and the derivatives were respect to the calendar time t and the current price of the underlying asset S_t. In contrast, Dupire's PDE is for the *initial* replication cost of a call option, and the derivatives are with respect to the maturity date T and the strike K.

REMARK. The point of the above theorem is this: Suppose you believe that the stock price is generated by a local volatility model, but you do not know what the local volatility function is. If you can observe today's call price surface $\{C_0(T,K) : T > 0, K > 0\}$ then you can solve for the local volatility in Dupire's PDE to arrive at *Dupire's formula*

$$\boxed{\sigma(T,K) = \left(\frac{2[\frac{\partial C_0}{\partial T}(T,K) + rK\frac{\partial C_0}{\partial K}(T,K)]}{K^2\frac{\partial^2 C_0}{\partial K^2}(T,K)}\right)^{1/2}.}$$

Furthermore, assuming Dupire's PDE has a unique solution (it will if σ is smooth and bounded as assumed) then we have found a model that can reproduce the observed call prices.

SKETCH OF PROOF OF DUPIRE'S FORMULA. 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, by the assumption of smoothness and the bounds on the volatility function, the $\mathbb{Q}$-law of the random variable S_T has a density function f_T. Computing expected values of both sides

$$(1) \qquad e^{rT}C_0(T,K) = (S_0 - K)^+ + \int_0^T \int_K^\infty f_t(y) y\ r\ dy\ dt + \frac{1}{2}\int_0^T f_t(K)K^2\sigma(t,K)^2 dt$$

and then differentiating both sides with respect to T yields

$$e^{rT}\left(\frac{\partial C_0}{\partial T}(T,K) + rC_0(T,K)\right) = \int_K^\infty f_T(y) y\ r\ dy + \frac{1}{2}f_T(K)K^2\sigma(T,K)^2.$$

[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*.

Now we use the following the Breeden–Litzenberger identities

$$
\begin{aligned}
e^{rT}C_0(T,K) &= \int_K^\infty f_T(y)y\ dy - K\int_K^\infty f_T(y)\ dy \\
e^{rT}\frac{\partial C_0}{\partial K}(T,K) &= -\int_K^\infty f_T(y)\ dy \\
e^{rT}\frac{\partial^2 C_0}{\partial K^2}(T,K) &= f_T(K)
\end{aligned}
$$

to finish the argument. □

CHAPTER 4

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*

$$\mathrm{Leb}(a, b] = b - a$$

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 $\mathrm{Cov}(X, Y) = 0$.

THEOREM. *Let X and Y be integrable random variables.*

- *linearity: $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ for constants a, b.*
- *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 either $X = 0$ almost surely or X and Y have the same sign and $|Y| = a|X|^{p-1}$ 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 $\mathrm{Cov}(X_i, X_j) = V_{ij}$.

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

$$\mathbb{P}(\bigcap_{i \in I} A_i) = \prod_{i \in I} \mathbb{P}(A_i)$$

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 $\{X_1 \leq t_1\}, \{X_2 \leq t_2\}, \ldots$ 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*

$$\mathbb{P}(X \geq \epsilon) \leq \frac{\mathbb{E}(X)}{\epsilon}$$

for all $\epsilon > 0$.

Corollary (Chebychev's inequality). *Let X be a random variable with $\mathbb{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$. Then*

$$\mathbb{P}(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

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*

$$\phi_X(t) = \phi_Y(t) \textit{ for all } t \in \mathbb{R}$$

if and only if

$$F_X(t) = F_Y(t) \textit{ for all } t \in \mathbb{R}.$$

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:*

$$\left.\begin{array}{c} X_n \to X \textit{ almost surely} \\ \textit{or} \\ X_n \to X \textit{ in } L^p, p \geq 1 \end{array}\right\} \Rightarrow X_n \to X \textit{ in probability } \Rightarrow X_n \to X \textit{ in distribution}$$

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 $\mathbb{P}(A_n$ *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 $\{1, 2, \ldots\}$
$\mathbb{C}$	the set of complex numbers
$\mathbb{Z}$	the set of integers $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$
$\mathbb{Z}_+$	the set of non-negative integers $\{0, 1, 2, \ldots\}$
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
$\mathrm{Var}(X)$	the variance of X
$\mathrm{Cov}(X, Y)$	the covariance of X and Y
$\mathbb{E}(X \mid B)$	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 $x_1, x_2, \ldots$
$\liminf_{n\uparrow\infty} x_n$	the limit inferior of the sequence $x_1, x_2, \ldots$
$a \cdot b$	Euclidean inner (or dot) product in $\mathbb{R}^n$, $a \cdot b = \sum_{i=1}^n a_i b_i$
$\lvert a \rvert$	Euclidean norm in $\mathbb{R}^n$, $\lvert a \rvert = (a \cdot a)^{1/2}$
$X \sim \nu$	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 $V \in \mathbb{R}^{n \times n}$
$\mathrm{bin}(n, p)$	the binomial distribution with parameters n and p
$\mathrm{unif}(a, b)$	the uniform distribution on the interval (a, b)
L^p	the set of random variables X with $\mathbb{E}\lvert X\rvert^p < \infty$

Table 1. Notation