Math finance, an intro to Option Pricing

Pietro Siorpaes

Contents

Week 1

1 What are derivatives?

1.1 Lecture 1, The Forward Contract

Derivatives (a.k.a. options[1]) are type of contract, which allows to transfer risk. To clarify what are derivatives, and what they are used for, it is best to start with some concrete examples.

The oldest-known derivative, already in use (to trade agricultural products) in India in 2000 BC and in ancient Babylonia in 1600 BC, is the *forward contract*, which we now introduce.

Example 1 (Forward Contract)*.* Today (i.e. at time 0) the airline easyJet buys from a bank a forward contract for A litres of jet fuel at *forward price* K (fixed today) and *expiry* (a.k.a. *expiry, maturity, delivery date*) T. This means that the companies enter into an agreement which states that, at the future time T, easyJet will buy from the bank A litres of jet fuel at price K.

Notice that a forward contract costs nothing, i.e. no cash changes hands at time today. Despite of this, we say that easyJet *buys* the forward contract, and the bank sells it, to indicate that at time T easyJet will buy the fuel from the bank. The forward price K is not the price which easyJet has to pay to buy a forward contract: this price, also called the value of the contract today, is zero. Instead, the forward contract is described by the parameters A, K, T, and the parameter K specifies the price (in some fixed currency, say £) at which easyJet will buy A litres of fuel from the bank at expiry T.

For easyJet, buying a forward contract removes the risk of rising fuel prices: if prices rise, easyJet will nonetheless have the right to buy the fuel from the bank at price K. This is particularly useful as easyJet sells tickets months in advance, so it cannot raise their prices to offset its increased costs due to the rise of fuel prices. However, if fuels prices drop, easyJet will also have the obligation to pay the pre-agreed price K at time T for the fuel, instead of paying its market price P_T. This is however an acceptable scenario: essentially, easyJet is sacrificing the possibility of a very positive outcome, in exchange for removing the risk of a very negative one. So easyJet, buying a forward contract, transfers the risk[2] of rising fuel prices to the bank. It can then focus on making money by operating its business well, without being subjected to the vagaries of the market.

In summary, the value of the forward contract at time T is $P_T - K$, since owning it allows, and forces, easyJet to buy at price K from the bank what it could have bought at price P_T from the market. Thus, the value F_T of a forward contract at time T depends

[1] Some people call *options* only a very special type of derivatives: those which we will call vanilla options.

[2] Notice how in finance the word 'risk' refers to any uncertain outcome, not necessarily to a negative uncertain outcome.

on (is a function of) the price P_T of the jet fuel at time T; more precisely,

$$F_T = f(P_T), \qquad \text{with } f(x) := x - K, \quad x \in \mathbb{R}.$$

The forward contract is then called a *derivative* (or a *contingent claim*), since it is a contract whose value derives from (is contingent on) the value that one or more other quantities, called the *underlying*, take at times $t \in [0, T]$.

The bank is now exposed to the risk of rising fuel prices. To reduce the risk, the bank will seek to trade with other customers in a way that the resulting risks cancel each other out as much as possible; this is called *hedging*. For example, Koch industries produces jet fuel and wants to cover itself against the risk of dropping fuel prices. Thus, the bank could buy the forward contract from Koch industries, and then sell it to easyJet. This trade reduces the risk for easyJet and Koch industries. In exchange for this service, the bank will make some small profits, because the forward price at which it will buy (*bid price*) and sell (*ask price*) it are actually slightly different.

Of course, Koch industries could have sold the forward contract directly to easyJet. However, in reality hedging is a lot more complicated, and one can never fully cover all risks. easyJet will not want to buy a contract with the same parameters at which Koch industries wants to sell it. Not even the bank, with its extensive connections to thousands of businesses, will ever be able to find someone from which to buy the same contract that it is selling to someone else. Instead, the bank will have to sell many types of derivatives to many participants, trading to hedge as much as possible the total risk resulting from its many investments. Koch industries and easyJet could also try to do that. However, to get the job done the bank has to employ innumerable highly paid math-whizzes to come up with a bewildering variety of sophisticated contracts, and figure out at what prices they should trade these, and find trading partners with which to conduct thousands of transactions in a way to minimize the combined risks, all the while satisfying complicated requirements from financial regulators. So, most companies prefer to concentrate their efforts on their core competencies, and pay a fee to the bank for its service. Large companies may however have their in-house financial arm; as an extreme example, reported by the Economist on 28/10/17 in the article Apple should shrink its finance arm before it goes bananas,

> [Apple's financial operation is] on some measures, roughly half the size of Goldman Sachs.

1.2 Lecture 2, The Binary Option

There are many possible derivatives and underlying, and a few more reasons to trade derivatives, as we will see. In this introductory chapter however we offer just one more example.

Example 2 (Binary option). Let's say it is 10am, gold is trading at \$58 per gram, and my American uncle Bob believes the price of gold will rise and close above \$59 by the end of the trading day (at 4pm EST). He can then place a wager on whether this will happen, asking his trader to buy for him from a bank 1500 *cash-or-nothing binary options* on gold with strike price \$59. Denote with B_0 the cost (in \$) of each binary option. Buying them, Bob enters a contract with the bank, which specifies that if

1. Bob's prediction is correct (i.e. at 4pm the price P_T of gold is above \$59), then the bank owes him \$100 for each option[3], i.e. \$150000 in total. The option is said to be *in the money*.

2. Bob's prediction is wrong, then the bank owes him nothing: Bob loses all the \$1500B_0 which he invested in buying the options. The option is said to be *out of the money*, as it is worth nothing.

In this example, the underlying is gold, the derivative is the binary option, and expiry is 4pm. Bob bought each contract at time 0 at price (i.e. initial value) \$$B_0$; its payoff at maturity (i.e. its final value) is \$100 if $P_T \geq 59$, and \$0 otherwise. The option's payoff can be written as $f(P_T)$ with $f(x) := 100 \cdot 1_{\{x \geq 59\}}$, where by definition

$$1_{\{x \geq b\}} := \begin{cases} 1 & \text{if } x \geq b, \\ 0 & \text{otherwise.} \end{cases}$$

The P&L (*Profit and Loss*) of Bob's investment in the binary option, defined as its payoff minus its cost, expressed in \$, is

$$\text{P\&L options} = \begin{cases} 150000 - 1500B_0 & \text{if } P_T \geq 59, \\ -1500B_0 & \text{otherwise.} \end{cases}$$

Remark 3 (Probabilities and prices)*.* This example shows how options allow traders to place bets of their market predictions; this makes it clear that the cost (initial value) of the option should be tied to the probability that their prediction comes true. More generally, the cost of a derivative will depend on the probability distribution (a.k.a. law) of the underlying, and potentially also of other trading instruments (tradable quantities).

1.3 Lecture 3, Speculation

Notice that Bob, convinced that the price of gold would rise, could have decided to buy gold instead of binary options. To compare this trading strategy with the binary option, assume to fix ideas that[4] $B_0 = 37$. If Bob invested the \$1500$B_0 = 55500$ in gold instead of in options, he would have bought $55500/58 \sim 957$ grams of gold, and then his final wealth would be $957 \cdot P_T$, i.e. his P&L would be $957 \cdot (P_T - 58)$. So, while both P&L's are strictly positive if $P_T \geq 59$, trading options can lead to much bigger gains and losses. For example, if $P_T = 59.3$ then

$$\text{P\&L options} = \$94500, \quad \text{P\&L gold} = \$957 \cdot 1.3 = \$1244.1,$$

[3] In the U.S., every binary option settles at \$100 or \$0: 100 if the bet is correct, 0 if it is not.
[4] As value that, as it will become clear later from our remark 8, is entirely reasonable.

whereas if $P_T = 57.7$ then

$$\text{P\&L options} = -\$55500, \quad \text{P\&L gold} = \$957 \cdot (-0.3) = -\$287.1.$$

Thus, trading with binary options is a risky move: its purpose is not to hedge away a risk, but rather to speculate. In fact, binary options are banned by regulators in many jurisdictions as a form of gambling! This example illustrates how derivatives can be used to create investments that are *riskier* than the underlying by 'magnifying' its price movements, making it a more effective way of betting on its future prices. This can be very profitable, but also very risky business, especially considering that trading in stocks is already quite risky: as Mark Twain splendidly put it in *Pudd'nhead Wilson's tale* (1894):

> October. This is one of the peculiarly dangerous months to speculate in stocks. The others are July, January, September, April, November, May, March, June, December, August and February.

As an horrific example (among many available) of the dangers of speculation, consider Barings Bank, which was one of England's oldest and most prestigious banks. It was founded in 1762 and collapsed in 1995, after an employee lost \$1.4 billion of company money speculating in futures contracts.

Remark 4 (Derivatives: good or bad?). Many people have very strong views -positive or negative- on derivatives, for several reasons:

1. derivatives can be used both for hedging and for speculating. While banks which trade derivatives and hedge their risks provide an important service which benefits other companies (and ultimately society as a whole), speculators can cause huge losses to their employers and customers, with disastrous knock-on effect for the rest of the economy.

2. the derivatives' market is huge.

3. derivatives are difficult to understand.

As a result, even experts on the topic can have a very wide range of views on derivatives. For example, while the legendary investor Warren Buffett wrote that

> [derivatives are] financial weapons of mass destruction [...] time bombs, both for the parties that deal in them and the economic system,

the Nobel laureate Merton Miller wrote[5]

> Contrary to the widely held perception, derivatives have made the world a safer place, not a more dangerous one. They have made it possible for firms and institutions to deal efficiently and cost effectively with risks and hazards that have plagued them for decades, if not for centuries.

Either way, what is undeniable is that derivatives have played a significant role in financial markets across the centuries, and that in today's complex world many institutions find derivatives essential in managing their varied risks.

[5] In 'Merton Miller on Derivatives'.

2 How to price and hedge derivatives in one-period models

2.1 Lecture 4, Pricing by replication

We will often use the word *portfolio*; by this we mean an investment, normally seen as a collection of investments in specific assets. For example, I could say that I have £10^6 in cash, 100 kg of gold and 10^4 shares of Apple in my portfolio.

Consider a derivative X which gives a payoff X_T at one[6] time in the future, denoted by T and called expiry. Assume that T is known at time 0, i.e. it is not random[7]. The derivatives payoff is[8] a known function of the present and past values of the underlying, i.e. it is of the form $f((P_t)_{t \leq T})$. In general, the bank needs to figure out at what price it 'should' sell the derivative today, i.e. what is its 'correct' initial value. The case of the forward contract is conceptually a little more complicated, as by construction the initial value of the forward contract must be zero; the bank here has to figure out the correct value of the parameter K such that the forward contract with final value $P_T - K$ has initial value zero; this value of K is called the *forward price* (of the underlying).

In many cases, the answer can be obtained simply by the following criterion:

Principle 5 (Law of one price)**.** *If there are two possible investments which have, under all possible market outcomes[9], the same value at time T, then they must have the same value also at all previous times.*

Later on we will show that, when an answer cannot be obtained by the law of one price, it can be obtained by a generalisation thereof (the domination principle, or equivalently the no-arbitrage principle). We will also eventually put our principles on a firm mathematical footing and justify why we should assume them. For now however, we will content ourselves with assuming the law of one price as self-evident, and concentrate on understanding how this principle can be used to price derivatives, using our two examples.

Consider again example 1: instead of buying the forward contract, easyJet could have decided to use the following trading strategy: buy the jet fuel today at price $\$P_0$, getting the money to finance this purchase by taking out a loan from a bank. Then at expiry easyJet would own the jet fuel, which has value $\$P_T$, and would owe the bank the amount \$L which was agreed[10] when signing the loan of $\$P_0$ at time 0; in other words, the final value of this investment is $\$P_T - L$. The value at time 0 was zero, since easyJet did not

[6] More generally, derivatives (e.g. swaps) can give a whole cashflow, i.e. a sequence of payoffs at multiple times $T_1 < \ldots, < T_n$. However, in discrete-time these derivatives can be seen as a portfolio of n derivatives, each with payoff at only one time $T = T_1, \ldots, T = T_n$. So, once we learn how to price derivatives which give a payoff at only one time, we can price derivatives with give a payoff at multiple times.

[7] We will eventually be able to apply the present pricing theory to options with random (non-anticipative) expiry T, but for now let us consider the simpler case of non-random T.

[8] Simply by definition of derivative and of underlying.

[9] Meaning, no matter what the value of the traded instruments turns out to be, among those values which are considered possible, i.e. which have a non-zero probability of happening.

[10] To lend money the bank will charge interest, and so $L = P_0(1 + r)$, where r is the interest rate.

use any of its money to buy the fuel (instead, it borrowed the money, so it has to repay it at time T). We can now figure out what should be the value of the forward price K: it should be L. Indeed, the above investment (borrowing from the bank to buy the jet fuel) has payoff $P_T - L$ and initial price zero. Since the forward contract with parameter L also has payoff $P_T - L$, then it should, by the law of one price, also have initial price zero; thus L is the forward price of jet fuel sold at time T. Since the above portfolio has the same payoff as the forward contract, we say it has *replicated* the forward contract. More generally, by the law of one price, if we can find a portfolio which replicates a derivative, we know that:

Principle 6 (Pricing via replication)**.** *At any time $t \in [0, T]$, the derivative's price must equal the value of the replicating portfolio.*

Of course there can be multiple replicating portfolios, but this leads to no contradiction, since (again by the law of one price) they will all have the same value at any time. In particular, the above principle implies that there is only one (reasonable) price at which to trade a replicable derivative

Remark 7 (Hedging = - Replicating)*.* Assume a trader sold a forward contract to easyJet. Instead of buying a forward contract from Koch industries it could, as described above, get a loan from a bank and to borrow just enough money to buy the A litres of jet fuel today at price $\$P_0$. She would then have exactly what it takes to fulfil her obligations from the forward contract, since she could deliver the A litres of jet fuel to easyJet in exchange for £K, and use those £K to repay her loan. More generally, the trader could sell a derivative, and use the proceeds from the sale to buy[11] a portfolio replicating the derivative; at maturity his obligations would be exactly matched by the value of his portfolio, thus covering his risks. This shows that replicating a derivative is the same as hedging 'minus the derivative', i.e. hedging having sold (one unit of) the derivative. It is then common to use interchangeably the terms *replicating strategy* and *hedging strategy*: though to be precise one it the negative of the other, it is always clear what one means.

In finance we say that the trader has gone short (/has shorted/has a short position in) the derivative if she has sold it, and has gone long[12] (/has a long position in) the derivative if she has bought it; with this language we can say that *replicating means hedging a short position*. By the way, in finance lingo to *write* a derivative means to sell it (and the *writer* of a derivatives means the seller).

Flipping around the above situation, a trader who bought the forward contract could decide to hedge her long position by taking the negative of the above trading strategy, i.e. selling[13] the A litres of jet fuel today at price $\$P_0$ and letting the proceeds accrue interest until maturity.

[11] He would have just enough money to do that, by principle 6.
[12] The expression 'has longed' is not used.
[13] As we will see later on (when discussing short-selling), the trader can sell the fuel even if she does not own it, essentially by borrowing it.

2.2 Lecture 5, Pricing a binary option

Let us now derive how the bank 'should' price the binary options in example 2. As mentioned in remark 3, to price a derivative in general we have to first model the law (probability distribution) of the available trading instruments. Let us assume for simplicity that in our market we can only:

1. trade gold, or
2. borrow/deposit money from/into a bank.

Also, since we are looking at a very short time span (between 10am and 4pm), let us assume for simplicity that the interest rate r is zero. We then only have to model the random behaviour of the price of gold at maturity P_T. Say for simplicity that we model it as a random variable which only takes the two possible values 61 and 56, both with probability $\frac{1}{2}$. So, we can think of a probability space $\Omega = \{g, b\}$ made of two possible outcomes ('g' for good and 'b' for bad) with probabilities $\mathbb{P}(\{g\}) = \frac{1}{2} = \mathbb{P}(\{b\})$, and take

$$P_T(\omega) = \begin{cases} 61 & \text{if } \omega = g \\ 56 & \text{if } \omega = b. \end{cases}$$

The payoff $B_T = 100 \cdot 1_{\{P_T \geq 59\}}$ of the binary option is then

$$B_T(\omega) = \begin{cases} 100 & \text{if } \omega = g \\ 0 & \text{if } \omega = b. \end{cases}$$

To price the binary option, let us try to replicate it. For added clarify, let us call 10am 'now' and 4pm time 'maturity'; all values are in \$. If we start with wealth $V_0 = x$ and we buy h grams of gold at price $P_0 = 58$ each, then at maturity our wealth will be

$$V_T = V_T^{x,h} = x - h \cdot 58 + h \cdot P_T.$$

This is true even if, to buy h grams of gold, we had to borrow \$$k$ from the bank, since in this case we would deposit \$$(x + k - h \cdot 58)$ in the bank now, and end up at maturity with \$$h \cdot P_T$ of gold and a debt of \$$k$ to the bank, and thus our final wealth would be

$$(x + k - h \cdot 58) + (h \cdot P_T - k) = x - h \cdot 58 + h \cdot P_T.$$

By definition, this is a replicating portfolio if its final wealth $V_T^{x,h}$ equals the payoff B_T of the binary option. This leads to the system of equations

$$\begin{cases} x + h \cdot (61 - 58) = 100 \\ x + h \cdot (56 - 58) = 0, \end{cases} \tag{1}$$

whose unique solution is $h = 20, x = 40$. Thus, there exists a portfolio which replicates the binary option: we should start with initial wealth 40 and buy 20 grams of gold. The initial value x of this portfolios 40, which by principle 6 must then be the price of the binary option.

Remark 8 (Replication depends on the model). When discussing example 1 it turned out not to be necessary to model the law of the underlying, because there exists a model-independent hedge, i.e. a trading strategy which replicates the forward contract *independently of the model*: buying A litres of jet fuel, borrowing the money to do so. In example 2 instead, the replicating portfolio depends on the model: if we changed 61 to 60, i.e. if we assumed that the P_T was given by

$$P_T(\omega) = \begin{cases} 60 & \text{if } \omega = g \\ 56 & \text{if } \omega = b, \end{cases}$$

then the replicating portfolio would have changed to $h = 25, x = 50$ (and thus the price of the binary option would be \$50.). Worse yet, if we assumed that P_T was modelled as possibly taking the three values $61, 57, 56$, we would have found that the binary option is not even replicable!

Week 2

2.3 Lecture 1, Model uncertainty

As remark 11 shows, it is extremely important to work with a market model that accurately describes reality. We will not discuss how to use statistics to do so, and instead refer the interested reader to [**?**].

Here is the crux of the problem. To price options we work in a fixed probabilistic model, e.g. we assumed P_T takes values 61 and 56, each with probability $\frac{1}{2}$. This is a situation where the realised outcomes are unknown, but the possible outcomes are governed by a probability distribution known at the outset. The famed[14] economist Frank Knight called this the 'known unknown', i.e. an unknown which is 'a quantity susceptible of measurement' (since for example we can calculate expectation, variance etc of such quantity). Given one such market model, we can find prices of options.

However, in reality, not only we do not know for sure which outcome will be realised: we do not even know the probability that govern the possible outcomes, so e.g. we could legitimately have modelled P_T as taking[15] values $60, 56$, or $61, 57, 56$, instead of $61, 56$. Thus, choosing a model introduces *model risk* (a.k.a. *model uncertainty*, or *Knightian uncertainty*, or *unknown unknown*), i.e. the unquantifiable risk that we have chosen an inappropriate model. This is the Achille's heel of option pricing; the 2008 financial crisis has been blamed on (among other things) the false security created by the over-reliance on models which 'disregard key factors', see e.g. the 13/05/2009 Wharton School article Why Economists Failed to Predict the Financial Crisis, which discusses the Dahlem Report. Let us now give two examples of how one could consider a more complicated but more realistic model than done do far.

Remark 9 (Storage Costs). Notice that we are ignoring for the moment the fact that easyJet, if it chose to replicate the forward contract (instead of buying one) by buying the fuel and borrowing the required capital, it would have to pay a fee \$F to store the jet fuel for 6 months. If our example substituted jet fuel with shares of a corporation (which are just contracts, and thus have no storage costs), or with gold (which has a small storage costs[16]), this would be perfectly reasonable. In other cases, e.g. when trading agricultural products (e.g. wheat, milk), it is probably never reasonable to ignore storage costs. In some cases, items just cannot be stored (at any cost) for more than a few days: for example flowers, since they wilt quickly. This, and the fact that roses are

[14] Frank Knight (1885-1972) had 3 students who became Nobel laureates, and was named one of the 'American saints in economics' by Paul Samuelson (Samuelson, who was the first of many Americans to win the 'Nobel in Economics', was considered by the New York Times to be the 'foremost academic economist of the 20th century').

[15] By which we mean that the probability that P_T equals any of those values is non-zero, and the probability that P_T equals none of those values is zero; e.g. we say that P_T takes values $61, 57, 56$ if

$$\mathbb{P}(\{P_T = x\}) > 0 \text{ for all } x \in \{61, 57, 56\}, \quad \mathbb{P}(\{P_T \notin \{61, 57, 56\}\}) = 0.$$

.

[16] Essentially the cost keeping it safe from thieves.

worth double during a very short period (Valentine's Day), has many consequences, as illustrated by the very interesting Episode 603: A Rose On Any Other Day of the (highly recommended!) Planet Money podcast. In other yet cases, like jet fuel or oil, ignoring storage costs is probably a good first approximation, though of course one always has to consider the specific market setting. For a remarkable example of how the market conditions matter, consider the following April 2020 article by the Guardian Oil prices dip below zero as producers forced to pay to dispose of excess, which details how the cost of crude oil became *negative* because of storage issues[17].

Remark 10 (Counter-party risk). Suppose easyJet buys the forward contract from Koch industries, jet fuel prices rise, and at maturity Koch industries is unable to sell jet fuel at the pre-agreed prices because it has gone bankrupt. Then easyJet's forward contract is worthless, and easyJet is decidedly not in the same situation as if had instead bought the jet fuel to replicate the forward contract. This shows that we have been ignoring *counter-party risk*, i.e. the risk that the other party in the trade will not fulfil its obligations. If we admit that the counter-party can[18] default, then the trading strategy we described is not actually a replicating strategy.

Remark 11 (Choosing the right model). As exemplified by remarks 9 and 10, one has to choose a model which is appropriate to the specific market one is considering. It is often not easy to do so, as it can be hard to realise what potential complications a given model is ignoring! While a model makes some assumptions, and is thus meant to be applied under specific market conditions, often its predictions will work reasonably well even if applied to a market which does not quite satisfy those assumptions/conditions: as George Box[19] quipped:

> Essentially all models are wrong, but some are useful.

Indeed, the prices which one gets from the no-arbitrage theory work so remarkably well in practice that the famed economist Stephen Ross said in 1987:

> [option pricing theory is] the most successful theory not only in finance, but in all of economics.

This is one of the main reasons why the derivatives' market has expanded massively since the beginning of the development of the no-arbitrage pricing theory (in the 1960's/1970's), and is now huge (about 8 times[20] the size the world's GDP :-o).

[17]Essentially, as the coronavirus pandemic depressed demand for crude oil way below the rate at which it was being extracted, and as the world was soon to run out of ways to store crude oil, oil producing companies became willing to pay to get rid of it (since the alternative was to temporarily close oil wells, which was more costly).

[18]i.e. that the event of default has a non-zero probability.

[19]British-born statistician, lived 1919-2013, he has been called 'one of the great statistical minds of the 20th century'.

[20]In 2017 the Global GDP was around $80 trillion in nominal terms, while the notional amount of the derivatives's market was about $636 trillion.

2.4 Lecture 2, What can be used as underlying

Let us now describe which assets[21] can be used as an underlying. Historically, derivatives were based on tangible physical goods[22], i.e. on

1. commodities[23] (e.g. (semi)-precious metals like zinc or gold, agricultural products like wheat or milk, crude oil, electricity, bandwidth);

nowadays instead most derivatives are based on financial assets[24] (so that, for simplicity, often people when talking of derivatives call the underlying simply 'the stock'), e.g.

2. bonds[25] (e.g. US Treasury bonds, UK gilts), interest rates,
3. stocks[26] (e.g. Amazon (AMZN), HSBC, Alibaba (BABA), Total (TOT)),
4. stock market indices[27] (e.g. SP500, NASDAQ)
5. FX[28] (exchange) rates (e.g. GBP/EUR, USD/RMB), including cryptocurrencies (e.g. Bitcoin).
6. other derivatives (e.g. a call option based on a call option, which is itself based on some other underlying).

An underlying can also be weather-related quantity (e.g. temperature, wind precipitation). However, the no-arbitrage pricing theory which we will develop cannot be applied to price weather-related quantities, since their underlying is not traded (though of course it is correlated with things[29] which are traded). Worse, as stated in [**?**, Chapter 1.4],

> 'a generally accepted framework for pricing temperature (or in general weather) derivatives does not exist'.

2.5 Lecture 3, The bank account and the bond

So far we said that our investor can 'put money in the bank', and receive an interest. While this is a conveniently intuitive way to talk about money markets, it is clearly not how things work in reality: what would it even mean for an investor to 'put money in the bank and get an interest', when the investor is a bank itself? In reality what does

[21] An asset is a resource with economic value.

[22] In economics, one calls *good* anything that satisfies a human want.

[23] Commodities are a type of goods whose quality may only differ slightly across its many possible suppliers; typically they are goods used as input in the production of other goods or services.

[24] i.e. on contracts.

[25] A bond is a contract that states that money is being borrowed (usually by governments, or large companies), and it to be repaid later with interest. As the borrower binds itself to repay, this agreement is called a bond.

[26] Shares are contracts which represent partial ownership of a company; all the shares of a company together are called a stock. So, one can say, e.g., one thousand shares of Tesla's stock.

[27] Indices are weighted averages of stock prices in a given geographical zone or industrial sector.

[28] FX=forex=Foreign Exchange.

[29] E.g. electricity consumption, power generated by wind and solar, agricultural production, etc..

exist is a *money-market*, in which one can trade *debt instruments*, i.e. contracts that describe how exactly the seller is borrowing money from the buyer and is to later repay at an interest. There is a whole variety of such contracts, with different names and characteristics, e.g. bonds (issued by national governments, corporations, municipalities etc.), debentures, notes, commercial paper, banker's acceptances, mortgage-backed securities, loans etc. As in this course we are not interested in the details of the money market, we will from now on simply talk of 'buying bonds', in the same simplistic way as we have talked about 'putting money in the bank', and we will always talk of 'the' interest rate, and use the following idealised but reasonable-enough representation of the money-market.

In other words, we will assume that we can invest in a contract, which we will call 'the bond' (or 'the bank account'), whose value B_t at time t is given as follows:

1. $B_0 = 1$, which we can assume w.l.o.g. by normalisation
2. if we work in discrete time: $B_{t+1} = B_t(1 + R_t)$, where R_t is the interest rate for investing in the time period between t and $t + 1$. Thus $B_t = B_0\Pi_{s=0}^{t-1}(1 + R_s)$, so if $R_t = r$ is constant $B_t = B_0(1 + r)^t$.
3. if we work in continuous time: B satisfies[30] $dB_t = B_tR_tdt$, where R_t is called the *instantaneous spot rate (a.k.a short rate)* at time t. Thus $B_t = B_0 \exp(\int_0^t R_sds)$, so if $R_t = r$ is constant $B_t = B_0e^{rt}$.

While sometimes we will model R as a stochastic process (as one should), often we will simply consider it as a constant, and write $R_t(\omega) = r$. We will *not* assume that $R_t > 0$ (i.e. we allow bonds to decrease in value), since the math does not require it and it is not always true[31]. We will however assume that $R_t > -1$, i.e. bonds always maintain some[32] value.

One should however keep in mind that in reality:

1. there is no such thing as 'the' interest rate: there are only many contracts, which specify how someone will borrow money from someone else during a set time period, and how she will repay the loan. There are several quantities that one could call 'interest rate', some of which are random quantities that changes in time (i.e. a stochastic process). These quantities also depend on the specific contract, e.g. in 2020 a bond issued by the German government will pay a lower interest rate than one issued by the Greek one (because lenders take into account the possibility that the borrower may default on its debt, and thus require a higher interest to compensate for the increased risk of default)

[30] This ODE (Ordinary Differential Equation) comes out by rewriting $B_{t+1} = B_t(1+R_t)$ as $B_{t+1} - B_t = B_tR_t$, so that changing the time step from 1 to $h > 0$ gives $B_{t+h} - B_t = B_tR_th$, and dividing times h and taking $h \to 0$ gives the ODE.

[31] As unintuitive as it may sound, interest rates have actually frequently been slightly negative between 2008 and 2020.

[32] While in reality governments and corporations can default, they do normally get to pay out at least a percentage of their debt in any case, so this is a reasonable assumption.

2. The market for debt is larger than the stock market, so studying money markets is very important. However, money markets are a complicated subject, so e.g. it is harder to price bonds and model interest rates than it is to price derivatives based on stocks and model stock prices.

2.6 Lecture 4, Justification of the law of one price

We have so far treated the law of one price as self-evident. Let us now justify it using the law of supply and demand[33]; we do this with the following fictitious example, which is a simplified version of what derivatives' trading is like.

Suppose that there are lots of Americans and Europeans living in Little Whinging. Some Americans want to exchange their \$ to €, and some the Europeans € to \$, but since they do not know one another, they have to go all the way to a bank in London to exchange currency. The Europeans can exchange € to £ using the exchange rate $S^{£}_{€} = 2$ (i.e. $€2 = £1$), and £ to \$ using the exchange rate $S^{\$}_{£} = 1.5$ (i.e. $£3 = \$2$); and the Americans the converse. Since going to the bank takes effort, an enterprising shop owner in Little Whinging turns into a trader, as she starts buying and selling to local customers contracts which state that at 10am tomorrow the buyer of the contract will give the seller €x and will receive from the seller \$1. Whoever buys or sells this contract to the trader also has to pay the trader a small fee, which most people prefer to do, rather than having to go through the trouble of going to the bank. Thus, Europeans start buying (*resp. Americans start selling*) contracts from (*resp. to*) the trader. Ideally the trader sells and buys exactly the same number of contracts, so that essentially Americans and Europeans switch their currencies using the trader as an intermediary, and the trader pockets the small fees. If however the trader ends up selling (*resp. buying*) more contracts that she buys (*resp. sells*), she can go early tomorrow morning to the London bank to buy the correct amount of \$ (*resp.* €) to satisfy the remaining obligation that she has to her customers (i.e she can always hedge her position...as long as $x = 3$). Suppose that in a 24h period the exchange rate between €,£ and \$ never changes significantly, and the interest rate is essentially 0. What value of x should the trader choose?

Well, since she can go to the bank and exchange €6 into £3 and then these into \$2, it seems very intuitive that she would set $x = 3$, which is indeed what the law of one price says; but what if she didn't?

If she chose $x = 2$, her sagacious customer Mr. Dursley would buy many (say M) of these contracts (for a fee so small we can disregard), go to London, borrow $£\frac{M}{2}$ from the bank, exchange them for €M, then use the contracts to get $\$\frac{M}{2}$ from the trader, then go to the London bank and exchange those $\$\frac{M}{2}$ to $£\frac{3M}{4}$; this way, Mr. Dursley gets to repay the loan to the bank, and make $£\frac{M}{4} > 0$ (the money made by Mr Dursley is of course lost by the customers who sell those same contracts to the trader; and by the trader, if she sells more contracts that she buys). Thus, *Mr. Dursley's investing strategy has made money starting with no initial capital to invest, and without any risk*; such an investment is called *an arbitrage.*

[33]Which we can use immediately as it is very intuitive; we will explain it in more detail later.

Now, if the trader kept $x = 2$, Mr. Dursley could buy an *arbitrarily big* amount amount M of contracts etc, and literally create a money pump that makes arbitrary amounts of money with no risk. However the trader, as Mr. Dursley buys more and more contracts to the point that the trader is not able to sell all of them, would raise the value of x, since this would make her customers want to sell more (and buy less) contracts, thus clearing again her inventory. This way, she is not exposed to the risk of not being able to hedge the obligations resulting from her growing inventory.

The trader would keep raising the value of x until $x = 3$, at which point Mr. Dursley's clever strategy doesn't work anymore (and there are no other arbitrage to strategies either). Analogously, if the trader chose $x > 3$, the law of demand and supply would quickly drive down the value of x to 3, at which point there are no more arbitrages. Thus, $x = 3$ is the only possible value which results in economic equilibrium; if at any time $x \neq 3$, some clever arbitrageur will quickly force x to revert back to 3.

Now, if you substitute:

1. the London bank for the (stock) market,
2. the trader for a bank (whose role here is that of a *marker-maker*, described later).
3. € and \$ for shares of two different stocks
4. the contract issued by the trader with any derivative based on those two stocks
5. the customers as any companies in need to buy or sell derivatives
6. Mr. Dursley with a bank employee working at the arbitrage trading desk

you essentially reproduced what happens when trading derivatives with two stocks as underlying.

Remark 12 (Transaction costs and arbitrage)*.* In the above, we have ignored the transaction cost incurred by the trader in trading with his customers. In practice, one cannot ignore these costs, as they may well be higher than the price imbalance detected by the arbitrageur in the market, in which case there is really no arbitrage!

2.7 Lecture 5, The no-arbitrage principle

The only substantial over-simplification of the above example is that the exchange rates are known in advance (as we assume they don't change over the 24h period). In reality, the value of the underlying is not known in advance, it can only be modelled as a stochastic process; correspondingly

1. the law of one price needs to be applied to random quantities (not just to deterministic payoffs, as it was in this example), as we already had supposed.
2. an *arbitrage* should not *always* make money. Rather, it should be defined as follows.

Definition 13 (Arbitrage)**.** An *arbitrage* is a portfolio (a.k.a. a *trading strategy*, or an *investing strategy*) L that, starting with no initial capital to invest, and without taking any risk, makes at money sometimes; i.e. an arbitrage is a portfolio L with zero initial capital and with final value V_T^L which satisfies $V_T^L \geq 0$ a.s., and $V_T^L > 0$ with non-zero probability (i.e. $\mathbb{P}(V_T^L < 0) = 0$, and $\mathbb{P}(V_T^L > 0) > 0$).

Nonetheless, this example:

1. conveys the fundamental idea that, if there ever was an arbitrage, the law of demand and supply would quickly change prices to make it disappear.
2. shows that what is really problematic is the existence of arbitrage, and that if the law of one price fails, then[34] there exists arbitrage.

Thus, whenever we model a market, we should choose a model that not only satisfies the law of one price, but more generally is arbitrage-free (i.e. admits no arbitrage). Indeed, an arbitrage is an unrealistic strategy, too good to be allowed to exist: while in the real world arbitrage fleetingly occurs, some traders (called arbitrageurs) quickly notice the arbitrage and take advantage of it and drive the prices back to equilibrium. Thus, we should exclude as non-sensical all models in which there exists an arbitrage.

Example 14 (Binomial model and arbitrage). Let us consider, in the one-period setting, the model of a market with only one risky asset S (which for simplicity we will call stock), plus the bond. So, we assume we can only invest at time 0 and receive some payoff at expiry T (normally one takes $T = 1$), and we can only invest in a bank account with interest rate[35] $r > -1$, and one stock whose value[36] $S_0 > 0$ is known at time zero (and it thus represented by a constant), and whose value S_T at time T is not known at time 0, and is thus represented by a random variable. If we now consider specifically the binomial model, we assume that S_T takes only two values, $S_0 u$ and $S_0 d$, with some probabilities $p \in (0, 1)$ and $1 - p$ respectively, where $u > d > 0$ (u is the *up factor*, d the *down factor*). When is this model arbitrage free?

To answer this question, assume from now on that our initial capital is 0. Then, to buy 1 share, we have to borrow £S_0 from the bank; if we do so, our final wealth is $W := S_1 - S_0(1 + r)$. If instead we buy $h \in \mathbb{R}$ shares (in which case we have to borrow[37] hS_0 from the bank), our wealth is hW. In particular, notice that if $h > 0$ then h is an arbitrage $\iff$ 1 is an arbitrage, since $\{hW \geq 0\} = \{W \geq 0\}$ and $\{hW > 0\} = \{W > 0\}$. Analogously, if $h < 0$ then h is an arbitrage $\iff -1$ is an arbitrage. Notice that $h = 0$ is never an arbitrage, since its final payoff is identically 0. So, *if we are considering a*

[34] We will see that this is true whenever we work in a linear model which admits a numeraire (i.e. essentially always).

[35] So, by definition, if we lend (/borrow) 1 to (from the bank at time 0, the bank owe us (/we owe the bank) $1 + r$ at time T.

[36] As measured in some currency, which we never specify unless we are dealing with two or more currencies at the same time

[37] Of course if $h < 0$ we are actually short-selling the shares (as discussed later), and correspondingly depositing $-hS_0 > 0$ in the bank.

model with only one risky asset, to check whether it has arbitrage or not it is enough to check whether $h = 1$ and $h = -1$ are arbitrages.

Taking $h = 1$, i.e. buying one share (while borrowing the corresponding amount of money), is an arbitrage iff $d \geq 1+r$, since in this case $W(T) \geq 0, W(H) > 0$. Analogously taking $h = -1$, i.e. selling one share (while depositing the corresponding earnings in the bank), is an arbitrage iff $1 + r \geq u$ since in this case $-W(H) \geq 0, -W(T) > 0$. Finally, if $d < 1 + r < u$ neither $h = 1$ nor $h = -1$ is an arbitrage, since the two corresponding final wealth W and $-W$ satisfy $W(H) > 0 > W(T)$ and so $-W(H) < 0 < -W(T)$, and so there exists no arbitrage. In conclusion:

$$\text{the binomial model is arbitrage-free} \iff d < 1 + r < u \tag{2}$$

2.8 Lecture 6, Dependence of prices on probabilities

To illustrate how prices depend on probabilities, let us revisit the example we considered in Section 2.2. Thus, consider a model in which Bob can only borrow/deposit money from/into a bank, with interest rate $r = 0$, and trade one asset (say gold, or shares of a corporation), whose price S Bob believes should be modelled as

$$S_T(\omega) = \begin{cases} 61 & \text{if } \omega = \omega_1 \\ 56 & \text{if } \omega \neq \omega_1, \end{cases} \qquad \text{with } \mathbb{P}(\{\omega_1\}) = 2/3.$$

In this case, Bob would price the binary option C with payoff $100 \cdot 1_{\{S_T \geq 59\}}$ by replication. To do so, he considers the system of equations

$$\begin{cases} x + h \cdot (61 - 58) = 100 \\ x + h \cdot (56 - 58) = 0, \end{cases}$$

whose unique solution is $h = 20, x = 40$. Thus, C can be replicated starting with an initial capital $x = 40$ and buying $h = 20$ units of the asset. So, the option should be sold at price 40, and thus Bob would be willing to sell it at any price ≥ 40.

Suppose that I believe instead that $\mathbb{P}(\{\omega_1\}) = 99.9\%$. At what price should I be willing to sell the binary option? To answer this question, notice that the probability $\mathbb{P}(\{\omega_1\})$ does not appear in the replication equation; thus $h = 20, x = 40$ is still a replicating strategy, and so I also believe that the option should be sold at 40. To understand why, consider that we are asking that the replication equation $V_1^{x,h} = C_1$ holds with probability one. Thus, if the probability $\mathbb{P}$ which describes Bob's beliefs of how market prices will evolve is replaced by the probability $\mathbb{Q}$ which describes my beliefs, as long as $\mathbb{P}$ and $\mathbb{Q}$ have the same null sets (sets of probability 0), the replication equation will be satisfied with probability 1 under $\mathbb{P}$ iff it is under $\mathbb{Q}$. Of course, in the above example one could consider as underlying probability space a set of two points $\{\omega_1, \omega_2\}$, and then the empty set is the only null set under $\mathbb{P}$, and also under $\mathbb{Q}$, which thus have the same null sets. It is thus convenient to introduce a notation: say that two probabilities $\mathbb{P}, \mathbb{Q}$ are *equivalent*, and write $\mathbb{P} \sim \mathbb{Q}$, if they have the same null sets.

Let us now see how prices change when switching to a non-equivalent probability. For that to happen (without having to assume that S_T is equal to a constant with probability one), we have to use as underlying probability space something bigger than $\Omega_2 := \{\omega_1, \omega_2\}$: let us use $\Omega_3 := \{\omega_1, \omega_2, \omega_3\}$. Of course, $\mathbb{P}$ can be considered as a probability on Ω_3, instead of Ω_2, by setting $\mathbb{P}(\{\omega_3\}) = 0$. Suppose now that Alice believes that

$$S_T(\omega) = \begin{cases} 61 & \text{if } \omega = \omega_1 \\ 56 & \text{if } \omega = \omega_2 \\ 55 & \text{if } \omega = \omega_3 \end{cases} \quad \text{with probability } \mathbb{P}'(\{\omega\}) = \begin{cases} 2/3 - \epsilon \\ 1/3 - \epsilon \\ 2\epsilon, \end{cases}$$

with ϵ very small, say $\epsilon = 10^{-9}$. Clearly $\mathbb{P} \not\sim \mathbb{P}'$, since $\mathbb{P}(\{\omega_3\}) = 0 < \mathbb{P}'(\{\omega_3\})$. To figure out at what price Alice is willing to sell the option, we cannot just use the law of one price, since in this setting the option is not replicable. One way to generalise the procedure of pricing by replication is to assume that Alice is infinitely risk-averse, i.e. that she does not want to take any risk, and ask at what price she is willing to sell the option. Clearly, the answer is at any price $\geq u(C)$, where $u(C)$ is the smallest super-replication price, i.e.

$$u(C) := \inf\{x : \exists h \text{ s.t. } V_1^{x,h} \geq C_T\}.$$

To be precise, the super-replication inequality $V_1^{x,h} \geq C_T$ needs to hold with probability 1, but as long as we model random variables as being defined on the space Ω_3, for which the only set with probability 0 under $\mathbb{P}'$ is the empty one, asking that the complement of the set $\{V_1^{x,h} \geq C_T\}$ is empty is the same as asking that it has probability 0 under $\mathbb{P}'$.

As we will show later on, if C is replicable then $u(C)$ is the same price as we calculated by replication, so indeed the above procedure generalises what we already know. Thus, $u(C)$ is the smallest x such that

$$\begin{cases} x + h \cdot (61 - 58) \geq 100 \\ x + h \cdot (56 - 58) \geq 0 \\ x + h \cdot (55 - 58) \geq 0. \end{cases} \tag{3}$$

To compute $u(C)$, we can draw the three half planes described by

$$\begin{cases} x + 3h \geq 100 \\ x - 2h \geq 0 \\ x - 3h \geq 0. \end{cases}$$

Since their intersection is the wedge between the two lines

$$x + 3h = 100, \quad x - 3h = 0,$$

to the right of their intersection at $x = 50, h = 50/3$, the smallest x such that (x, h) belongs to the above three half planes for some h is $u(C) = 50$. Notice that, since the model depends on ϵ, so in principle does $u(C)$, which we could thus denote with $u^\epsilon(C)$; however, as the value of ϵ does not appear in the system eq. (3), so $u^\epsilon(C) = 50$ will not actually depend on ϵ...as long as $\epsilon > 0$. If instead $\epsilon = 0$, we are back to the case of Bob, and thus $u^0(C) = 40$.

This shows that the 'price' $u^\epsilon(C)$ changes drastically as soon as ϵ goes from 0 to any value > 0. If ϵ is very small, Alice beliefs are extremely similar to those of Bob, and yet the price at which she is willing to sell the option is wildly different from that of Bob. This is **not** really reasonable, since in reality it is impossible to (super-)replicate anything with absolute certainty; more generally, the outputs of a good model should depend continuously on the inputs, since one can never know anything (neither the 'right' model, nor the value of the inputs) with absolute precision. Even if this is not reasonable, that is how no-arbitrage pricing works, so get used to it: it works well enough in practice, often enough, and that's what matters. It is good to keep in mind however, that you should never blindly trust the output of any model.

2.9 Lecture 7, Unspoken modelling assumptions

When considering example 14, on top of the assumptions that there is only one time period and only one stock price which can only take two values (of course these assumptions are ridiculous, but we will improve on these), we made some very common modelling assumptions, which we will essentially make throughout this course and which, as it is so often the case, went unnoticed and unchallenged. Let us critically discuss them.

Remark 15 (Information). In all our considerations we assumed that all market participants have the same information (there are no 'insiders'); the theory otherwise changes completely. This assumption is of course blatantly untrue, though it is not too unreasonable to assume that all major financial institutions have mostly the same information available. A nice example of of how to take advantage of additional information is offered at the end of the famous 1983 movie Trading Places; for an explanation thereof, one can listen to Episode 471 of the Planet Money Podcast.

Remark 16 (Linear dependence). We took as $h \in \mathbb{R}$ the number of shares we purchased, and we assumed that 'buying' $h \in \mathbb{R}$ shares at time 0 would 'cost' us hS_0 (i.e. buying $h > 0$ would cost $hS_0 > 0$, and selling $h > 0$ would yield $hS_0 > 0$), and 'buying' $k \in \mathbb{R}$ bonds would cost us kB_0. In other words, we assumed that the portfolio (k, h), which holds k bonds and h shares, has value $V_s^{k,h} := kB_s + hS_s$ at time $s \in \{0, 1\}$.

More generally, when working in discrete-time[38], and whatever model for the bond and shares prices B, S we choose, if the stochastic processes K, H represent the number of bonds and shares which we hold between time t and time $t + 1$ (and H is a vector valued quantity $H = (H^1, \dots, H^m)$, as H^j represents the number of shares in the stock of type j, whose price is S^j), then we take

$$V_s^{(K,H)} := K_t B_s + H_t \cdot S_s, \text{ for } s = t, t + 1 \tag{4}$$

as the value of the portfolio (K, H), where $S = (S^1, \dots, S^m)$ are the stock prices and $\cdot$ is the dot product on $\mathbb{R}^m$. In other words, we normally assume that the dependence

[38] Eventually we will even see how to generalise eq. (4) to continuous time, and obtain a formula for the value $V_t^{(K,H)}$ at time t of the portfolio (K, H).

$(K, H) \mapsto V^{(K,H)}$ is linear. As reasonable as this may sound, as anyone with any experience modelling knows:

1. if anything behaves linearly, in reality it only does so on first approximation.
2. linear models are infinitely easier to work with than non-linear ones, and are normally the only models that one can solve. Thus, reasonable or not, that's really the only choice we have.

Thus, you should not be so surprised to know that our linearity assumption is not really satisfied (but it is absolutely needed). In particular, in reality:

1. one can only buy or sell an integer number of shares. This is however not a problem, since each share is worth very[39] little (so, if you are buying 15000 instead of 15000.1 shares, it really doesn't make much of a difference).
2. the prices at which a share can be bought and sold (called bid and ask price) are not quite equal. When (as it if often the case for liquid shares) the difference between these numbers is small, we can reasonably ignore this detail.
3. the interest rate for lending and borrowing cash are not quite equal. However, for banks and other (large and reputable enough) investors, the difference between these numbers is small, so we can reasonably ignore this detail.
4. if buying one share of a stock costs S_0, buying a big number h of shares costs *more* than hS_0, since by the law of demand and supply an increase in buy orders will trigger an increase in prices, and so the shares we buy later will cost more than the first ones we buy (correspondingly selling $h > 1$ shares one makes less than hS_0). More precisely, the above reasoning shows that the cost $c(h)$ of buying h shares of a given stock is a function whose derivative $\frac{dc}{dh}$ is not constant, but is instead strictly increasing in h; consequently, the map $h \mapsto c(h)$ is not linear, rather, it is strictly convex. Of course, as long as h is not too big, i.e. if we make the assumption that our trader is too small a participant in the market to affect prices by its actions, this effect is also very small and can be ignored (essentially because we are approximating the cost function c by its tangent at 0).
5. a trader cannot choose to hold whatever amount h of shares she wants, because the total number of shares is finite and there might be legal or practical complications in buying too many shares. Analogously, the trader cannot borrow an arbitrarily large amount of money from the bank. These considerations also do not cause problems if we assume that our trader only trades 'little'.
6. buying or selling involves 'market frictions' like transaction costs, taxes, the cost of running the business (paying for employees) etc. These frictions cannot always be ignored, but considering them normally makes the models non-linear and thus essentially intractable.

[39] Normally up to a few thousand £/\$/€ each; for financial investors, this is so little that normally shares are bought in multiples of 100.

7. we assumed that a trader can sell $h > 0$ shares which she does not own. This practice, called *short-selling*, deserves to be discussed in a separate remark, as we now do.

Remark 17 (Short-selling). In example 14, if $h > 0$ the trader uses hS_0 units of cash which she does not own to buy one share and make an arbitrage; this she can do by borrowing $hS_0 > 0$ from the bank, and returning it with interest at time 1.

Analogously (though maybe less intuitively), if $h < 0$ the trader can use one share *which she does not own*, sell it and with it buys some cash and makes an arbitrage. Selling a share which you do not own is called *short-selling*, and (just like with cash), the above strategy in practice involves borrowing[40] the share from a bank or a broker, selling it, and then at a later time buying it from the market and returning it to the lender. While in our models we ignore these complications, in reality short-selling is really quite problematic and does not occur that frequently, since

1. it is really risky, as one has to buy the share at a later time, and while the price of a share can only go down to 0 (so there is a limit to how much one can lose by buying shares), it can go up to any arbitrarily high value

2. one has to pay a fee for the privilege of borrowing, making it ever harder and more risky to make money by short-selling

3. short-selling shares of a company is essentially placing a bet that the company's value will decrease, and so the short-seller will try to convince enough people to emulate his/her strategy, so that by the law of demand and supply the value of the company's shares will indeed decrease and perhaps crash. Thus, companies and even countries hate[41] short-sellers, and fight them as they can. In particular, short-selling is sometimes temporarily out-lawed by a government, or banned by some trading platforms, especially during financial crisis.

Thus, as long as we assume that we are considering a 'small trader', treating short-selling in a manner that is completely analogous with buying with just a minus sign in front, and ignoring market-frictions, are the only items listed in remark 16 which are not very realistic.

It is quite humbling to see how many things can go wrong with our model, despite it looking quite reasonable, isn't it?

Remark 18. Some of the above complications can be considered by replacing our linear model with a linear model with some (additional) linear constraints; these are also nice and solvable, though they do nonetheless involve some additional complications.

[40] A few market participants (e.g. market makers) may be allowed to short-sell even shares that they have not borrowed in advance.

[41] See e.g. the 23/06/2018 Bloomberg article Why Short Selling Can Make You Rich But Not Popular and the CBC 19/09/08 article Why they hate short sellers.

Week 3

2.10 Lecture 1, Prices of liquid and illiquid goods

We have showed how to compute the price of derivatives based on an underlying which is given a by some (e.g. binomial) model. At this point, the reader might be puzzled and have some legitimate questions, such as:

1. What determines the price of the underlying ?
2. Why did it make sense to model the price of the underlying (using random variables)?

To figure out what determines the price of the underlying, we need to talk about the law of demand and supply, which is what economists normally use to determine prices. Suppose that the demand and supply of some commodity, say fuel, are equal, so its price is stable. Suppose then that something causes the demand for fuel to rise (e.g. many more people buy cars). Then the fuel 'sellers' (by which we mean the aggregate of its whole supply chain: all those who pump oil out of the ground, transport it across the world, refine it into fuel, and then sell the fuel to car drivers at the gas pump) would raise the fuel price, since even doing so they'd still be able to sell their supply. The increase in price would cause some fuel buyers to consume less fuel (e.g. by replacing car driving with public transport), and some fuel sellers to produce more fuel (e.g. by opening new oil wells which, due to their geology, cost more to operate and thus only stay open when oil prices are high enough). This would thus reduce the demand and increase the supply, and it would keep on happening until demand = supply. The converse would also happen: if something causes demand for fuel to decrease (e.g. a Covid-induced lockdown), so that demand becomes lower than supply, then prices would fall, making the demand rise and lowering the supply until demand = supply again. So, the prices of commodities, shares and the other assets which we can use as underlying, arise from the interactions between the different market participants (as described by the law of demand and supply), and thus we say that they are 'priced by the market'.

Notice that the law of demand and supply has extremely broad applicability, e.g. you can apply it to determine the prices of commodities, capital, labour, etc; unfortunately, it does not lead to solvable, tractable models, other than toys models which are much too simple and unrealistic to be of value. Thus, the law of demand and supply, while offering a nice and very general theoretical justification of a pricing mechanism, can only provide a qualitative description of prices, not a quantitative one. Since we cannot use the law of demand and supply to build a reasonable model which outputs a stochastic description of the evolution of prices of commodities, shares etc., we instead model these prices directly.

While the law of demand and supply is broadly applicable, it cannot be used to determine price of *illiquid (i.e. non-traded) goods.* As an extreme example, suppose the Italian state was in such dire financial straits that it decided to sell the Colosseum. At what price should it be sold? Surely we cannot answer such a question by looking at existing markets to come up with a reasonable price; instead, here one would have to set up an auction. So, it is important to distinguish between illiquid markets, and liquid ones; the latter are those in which a good is traded a lot, by many possible market participants. The law of one price can be used to price liquid goods, but not illiquid ones.

We will always consider two extreme and idealised situations: markets which are either (completely) liquid or (completely) illiquid; real markets are of course always somewhere in the middle. For example, while call options far out of the money (i.e. for which the strike price K is very different from the initial value S_0 of the underlying) are illiquid, call options at the money (i.e. for which $K = S_0$) are liquid. Between these two extremes, we have the whole spectrum, so call options near the money (i.e. for which $K \approx S_0$) are a bit less liquid than those at the money, and as K gets further and further away from S_0 the corresponding call options become less and less liquid.

Goods that are normally very liquid are financial assets such as bonds, stocks of major companies (after their IPO[42]), stock indexes, and FX, which is the largest and most liquid market of all. Other goods that are normally liquid are stocks in smaller companies, and commodities. Some derivatives' market are also liquid: for example, futures, and near-the-money call and put options, based on (i.e. whose underlying is) a major currency or a major stock market index.

Good that are quite illiquid are: real estate, luxury items (e.g. antiques, art pieces, Ferraris), heavy machinery (e.g. industrial equipment, battle tanks) and, most importantly for us, most derivatives: especially those based on a not very liquid underlying (like stocks of small companies), or with a expiration date far into the future.

Notice that liquid derivatives, as all liquid goods, are priced by the market via the law of demand and supply (and thus their prices should be modelled). Traders don't need to figure out at what price to sell liquid derivatives, nor any liquid goods: the market does it for them. Instead, option pricing deals with how to price illiquid derivatives based on a liquid underlying, using the law of one price and its generalisation (the Domination Principle, i.e. the no-arbitrage principle). This uses a different pricing mechanism, which has a much narrower field of applicability, but which allows to consider, and deal with, complicated models and obtain useful quantitative results from them (sometimes analytically, sometimes just numerically).

2.11 Lecture 2, Modelling the underlying, not the derivative.

While traders can observe in the market the present-time value of liquid goods, they don't know how market forces will shape their future prices. However, simple observation about their historical behaviour teaches them that prices vary in an unpredictable, wildly erratic manner: for example, here a chart of coffee prices: Though this differs so drastically from the smooth dependence that one normally encounters in classical physics, this behaviour is to be expected: it is the reason why traders cannot be sure

[42] IPO=Initial Public offering; before the IPO, there is no market for a stock.

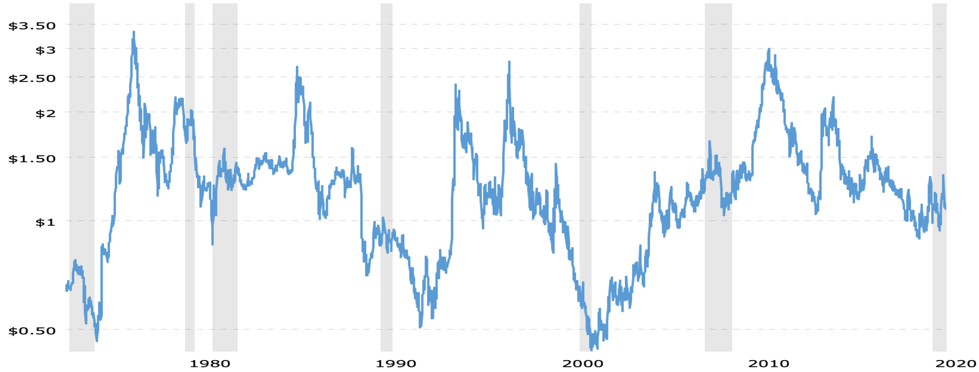


Figure 1: Image courtesy of Macrotrends

of what is the best way to invest, and thus choose to 'insure' themselves against price uncertainty by hedging away the risk of losses using derivatives. Indeed, if prices were differentiable in time, it would be extremely easy to choose how to invest, since the behaviour of prices in the recent past would be a near perfect approximation of their behaviour in the near future, so one would simply have to invest in the stock that had the best performance in the recent past!

Moreover, this erratic behaviour could perhaps be expected from the law of demand and supply. Indeed, at any time an order is issued, prices will jump, because of the sudden imbalance between supply and demand; if we consider a market where some orders are *large*, then the corresponding jumps in prices will be big, and we are forced to model these prices using a stochastic process with jumps. Suppose instead that the market for some good is made of frequent[43] *small* transactions (e.g. if we assume that all traders are 'small'). Then its price will move by lots of little jumps, which from a macroscopic point of view can be looked at as a path which is continuous, but is very jagged and oscillates wildly. Thus, in this (continuous-time) setting it makes sense to model price with a stochastic process whose paths are continuous, but very rough (e.g. not of finite variation). This in turn creates some mathematical difficulties, which have been solved by the development of the (hard!) theory of stochastic integration, which considers measures which have values in the space $L^0(\mathbb{P})$ of all random variables on a given probability space (instead of in $\mathbb{R}$), or a subspace thereof.

To all of us living in the 21st century, as we have all been exposed to the rudiments of probability theory, it seems pretty obvious that a quantity whose future behaviour is not known should be modelled mathematically as a *stochastic process*, i.e. random variable that changes over time (more precisely, a family $Y = (Y_t)_{t\in[0,T]}$ of random variables $Y_t = Y_t(\omega)$, parametrised by time $t \in [0,T]$). But as probability theory burst on the scene very late[44], it is absolutely remarkable that in 1900 the French mathematician and former stock trader Louis Bachelier, *in his PhD thesis*, proposed to model stock prices as a stochastic process (Brownian Motion $W = (W_t)_{t\geq 0}$) which he 'invented' (beating

[43]If the market for a good is liquid, transactions will be frequent.
[44]The definition of probability space, due to A. Kolmogorov, only appeared in 1933.

Einstein to the punch by 5 years), and whose paths (i.e. the functions $t \mapsto W_t(\omega)$) display exactly the type of erratic behaviour expected of prices.

Though Bachelier realised that, as Brownian Motion took also negative values, it did not make perfect sense to use it to model stock prices, it was only half a century later that Paul Samuelson, after being introduced to Bachelier's work (which, having been way ahead of its time, had been ignored and forgotten, until the statistician Jimmy Savage discovered it and advertised it in the 1950'), decided to consider instead Brownian Motion on an exponential scale, i.e. to model the stock price as $S_t = \exp(\sigma W_t + ct)$ (here $\sigma, c > 0$ are constants). This is the celebrated model named after Black and Scholes, who managed to compute the price of a call option in such model; Scholes and Merton won the Nobel prize in 1997 (Black having already died) for this and related work. We refer to [**?**] for more historical information on the topic.

While the Black and Scholes model has several shortcomings, it remains the benchmark model against which to compare. Bachelier's idea to treat stock prices as a 'known unknown', i.e. to model their price as a stochastic process, stands today. Traders select a family of processes $(P^\lambda)_\lambda$, which depends on some parameter[45] λ whose value is to be chosen via statistical considerations, to make the corresponding model best fit the observed behaviour of the market. Thus, once the parameter $\lambda = \lambda_*$ has been chosen, they describe the prices of a given liquid good with a stochastic process $P := P^{\lambda_*} = (Y_t^{\lambda_*})_{t\in[0,T]}$. The price at time t is given by the random variable P_t, whose value becomes known[46] at time t. In particular, the price P_0 of a liquidly traded good is known already at time 0, i.e. it is deterministic (i.e. it is a constant 'random' variable).

2.12 Lecture 3, Discounting and Numeraire

We now introduce a useful accounting practice, named *discounting*. As the saying goes,

> a dollar today is worth more than a dollar tomorrow.

In other words, normally one unit of currency (hereafter we consider \$, to fix ideas) decreases in value, if we measure value by the amount of (most) goods one \$ can buy, so e.g. 100 years ago \$1 was worth about 13 of today's \$ (see US inflation calculator). This is due to the fact that normally people are only willing to take the risk (and the loss of opportunity) that comes with lending money if they are to receive back their loan with interest; historically this has almost[47] always been true.

Thus, when comparing values across times, it make little sense to use currency (e.g. \$, £, €, ...) as a unit of measure. Rather, one should use the value B of 'the bond'; in other words, instead of looking at the value W_t of something in \$, we should instead consider the value

$$\overline{W}_t := W_t/B_t$$

[45]The parameter can be a real number, a vector, a function, or even a stochastic process.

[46]We will develop the mathematics to make sense of this

[47]There have been some (few) exceptions, e.g., since the 2008 financial crisis central banks often kept interest rates slightly negative; in medieval times Cristians were prohibited from lending money at an interest by the (Roman Catholic) church, etc.

in units of B. Often people call W the price in *nominal terms*, and $\overline{W}$ the price in *real terms* (since it better reflects how many things how can buy with that value), or the value of W *adjusted for inflation*, or the *present*[48] *value* of W. We will normally work in discounted terms determine $\overline{V}_t$; if we ever need to compute the value V_t in in \$, we can then simply multiply times B, since $V_t = \overline{V}_t B_t$.

Consider from the moment the one-period linear model of eq. (4), i.e. given constants $B_0 = 1, B_1 = B_0(1+r) > 0, r > -1, S_0 \in \mathbb{R}^m$ and the random vector $S_1 \in \mathbb{R}^m$, consider the portfolio (k, h) with value $V_s^{k,h} := kB_s + h \cdot S_s$ at time $s \in \{0, 1\}$. Since the initial capital is a quantity more of interest than the investment in bonds, we will normally change variables, and describe the portfolio as (x, h), where $x = k + S_0 h$ is the initial capital and h^j the number of shares of stock S^j (where $j = 1, \ldots, m$). The value of (x, h) is

$$V_0^{x,h} := x, \quad V_1^{x,h} := x(1+r) + h \cdot (S_1 - S_0(1+r)), \quad (x, h) \in \mathbb{R} \times \mathbb{R}^m. \tag{5}$$

Since $\overline{V}_t^{x,h} = V_t^{x,h} / B_t$ we find that $\overline{V}_0^{x,h} = V_0^{x,h} = x$, and the formula eq. (5) for the value of a portfolio in our linear model, when expressed in discounted terms becomes

$$\overline{V}_1^{x,h} = \frac{1}{1+r} \left(x(1+r) + h \cdot (S_1 - S_0(1+r)) \right) = x + h \cdot \left(\frac{S_1}{1+r} - S_0 \right),$$

and so in summary

$$\overline{V}_t^{x,h} = x + h \cdot (\overline{S}_t - \overline{S}_0), \quad t = 0, 1. \tag{6}$$

Remark 19. Working in discounted terms is certainly an intuitive way to describe values; it also has several additional advantages:

1. Since $\overline{B}_t = 1$, when working in discounted terms the interest rate r is always 0; conversely, if the interest rate is zero then values in nominal and real terms coincide, so even if we were working with nominal values, we could pretend that were working with real values. In other words, discounting can be seen as an accounting trick which allows us to assume w.l.o.g. that $r = 0$. This be can useful and it can simplify proofs and calculations.

2. The full usefulness of discounting will only become clear in the multi-period (and in the continuous-time) settings, where it will allow us to automatically take care of the self-financing condition.

3. In discounted terms, the *gains from trade* between times 0 and t are given by

$$\overline{V}_t^{x,h} - \overline{V}_0^{x,h} = h \cdot (\overline{S}_t - \overline{S}_0); \tag{7}$$

[48] It is called *present value* because to replicate the amount $\$x$ at a future time T, one could simply deposit $\bar{x} := x/B_T$ in the bank at time 0 (time 0 being though of as 'the present'). So, $\bar{x}$ represents the value at time 0 of receiving x at time T

it is important that this expression involves multiplying *the increment of something* (the discounted price $\overline{S}$) times something else (the trading strategy h). Indeed, in the continuous-time setting we will have look at the gains in the infinitesimal interval between t and $t + \epsilon$ as $\epsilon \downarrow 0$, and the above expression will become

$$d\overline{V}_t^{x,H} = H_t \cdot d\overline{S}_t, \quad \text{or equivalently } \overline{V}_s^{x,H} - \overline{V}_0^{x,H} = \int_0^s H_t \cdot d\overline{S}_t,$$

and so we will be able to express the quantity of interest (the discounted value $\overline{V}^H$ of the portfolio H, and thus also its nominal value $V^H = B\overline{V}^H$) in some way (by integrating[49] with respect to $\overline{S}$). It is not clear how we could have generalised to the continuous-time setting the expression eq. (5) for the nominal value $V_t^{x,h}$.

More generally, if V^L if the value in \$ of some portfolio L s.t. $V^L > 0$, if something has value W_t in \$, we can express its value as W_t/V_t^L at time t in units of V^L. This motivates the following definition.

Definition 20 (Numeraire)**.** A *numeraire* is a portfolio L which has a strictly positive value a.s. and at all times, i.e. s.t. $V_t^L > 0$ a.s. for all t.

Since in our linear model we assumed that the interest rate satisfies $R_t > -1$, the bond is a numeraire. In theory an investment in shares of a specific company cannot be a numeraire, since any company has a non-zero probability of going bankrupt at some point; whereas an investment in gold or[50] in a[51] major currency can be considered as a numeraire. In practice however, one normally works with models where the stock price is assumed to be strictly positive, and one can then use it a numeraire.

Remark 21. While we normally use the bond as a numeraire, it can be useful to consider other numeraires. This is most intuitive when one is considering a problem that involves multiples currencies, in which 'the bond' is replaced by multiple bonds (one for each currency), whose interest rates are different and whose ratio changes over time (and is random); in this setting it would make sense for an American investor to use 'the US bond' as a numeraire, and for a British investor to use 'the UK bond' instead. Moreover, often choosing an appropriate numeraire can simplify the calculations. It can also reduce the calculations of the price of a derivative to those for the price of a different derivative. E.g. using the asset S^1 as numeraire, the derivative with payoff $S_T^1 f(S_T)$ (where $S = (S^1, \dots, S^m)$) will have discounted payoff $f(S_T)$, of which we can try to find the value by working in the discounted market $(\tilde{B}, \tilde{S})$, where $\tilde{X} := X/S^1$. Analogously, when considering the bond as numeraire and the derivative with payoff

[49] Here $\int_0^s H_t \cdot d\overline{S}_t$ is not the usual Lebesgue-Stieltjes integral, but a more complicated object, called (vector) stochastic integral.

[50] For a long time, when the world was on the gold standard, the value of the £ (and later of \$ etc) was *defined* using the value of gold, i.e the price of gold in £ was constant in time; in that word, using £ as a numeraire was the same as using gold (up to a constant).

[51] While even major currencies can cease to exist (e.g. Deutsche mark), they do not do that suddenly, so one can change its investment into another currency and thus keep having a positive value.

$B_T f(S_T)$, we end up with the discounted payoff $f(S_T)$, of which we can try to find the value by working in the discounted market $(\overline{B}, \overline{S})$, where $\overline{X} := X/B$. Thus, by changing market (from $(\tilde{B}, \tilde{S})$ to $(\overline{B}, \overline{S})$), one can reduce the calculation of the price of the first derivative to the price of the second derivative, in the sense that one has to deal with the same discounted payoff $f(S_T)$ in both cases.

Remark 22. Clearly which numeraire we use to perform the calculations will not change the properties of our models, so e.g. whether a strategy is an arbitrage, or a derivative is replicable, does not depend on the numeraire used for calculations; the arbitrage-free prices of a derivative can be calculated using any numeraire and then converted to another, etc..

2.13 Lecture 4, Finite Probability Spaces

Given a finite set $\Omega = \{\omega_i : i = 1, \dots, n\}$, we can consider the σ-algebra $\mathcal{A}$ of its parts (i.e. the family all its subsets $\mathcal{A} := \{A | A \subseteq \Omega\}$), and endow $(\Omega, \mathcal{A})$ with some probability $\mathbb{P}$. W.l.o.g. we can (and will) assume that $\mathbb{P}(\{\omega_i\}) > 0$ for all $i = 1, \dots, n$, since otherwise we can simply remove from the space Ω all such points as they are irrelevant, i.e. we can replace Ω with $\Omega' := \{\omega_i \in \Omega : \mathbb{P}(\{\omega_i\}) > 0\}$ and $\mathcal{A}$ with $\mathcal{A}' := \{A | A \subseteq \Omega'\}$, and $\mathbb{P}$ with its restriction $\mathbb{P}'$ to $\mathcal{A}'$. We will call a triple $(\Omega, \mathcal{A}, \mathbb{P})$ as above a finite probability space. When working with it, we will use the very convenient practice of representing a random variable X on Ω by the vector $x = (x_i)_i \in \mathbb{R}^n$ whose i^{th} component is $x_i = X(\omega_i)$. Analogously, we will represent the random vector $X = (X^j)_{j=1}^m$ with the matrix $x = (x_{i,j})_{i,j}$ given by $x_{i,j} := X^j(\omega_i)$ (here i enumerates the rows and j the columns, so that a random variable is represented by a column vector). Analogously a measure[52] $\mathbb{P}$ on $(\Omega, \mathcal{A})$ will be represented by the arbitrary vector $p = (p_i)_i \in \mathbb{R}^n$ through the following identification

$$p_i = \mathbb{P}(\omega_i) := \mathbb{P}(\{\omega_i\}) \qquad \mathbb{P}(A) = \sum_{i:\omega_i \in A} p_i, \quad \text{for all } A \subseteq \Omega$$

and $\mathbb{P}$ is a probability (i.e. a positive measure of norm[53] 1) iff p is a positive vector (i.e. $p \in \mathbb{R}^n_+$) such that $\sum_i p_1 = 1$. Keeping the above identifications in mind, instead of pedantically distinguishing between X and x, $\mathbb{P}$ and p, we will normally simply write X to mean x, and $\mathbb{P}$ to mean p; whether X (*resp.* $\mathbb{P}$) should be considered as a random variable (*resp. a probability*) or a vector will become clear from the context.

Thanks to our assumption of linearity, if we work in a finite probability space, all pricing and hedging problems in our model of eq. (5) are reduced to questions about systems of finitely-many linear equalities and inequalities (from an algebraic point of view), i.e. questions about finite-dimensional vector spaces and polyhedra[54] (from a

[52] i.e. a function $\mathbb{P} : \mathcal{A} \to \mathbb{R}$ which is σ-additive, i.e. s.t. $\mathbb{P}(\cup_n A_n) = \sum_n \mathbb{P}(A_n)$ for every sequence of disjoint sets $A_n \in \mathcal{A}, n \in \mathbb{N}$.

[53] Using the total variation norm.

[54] By definition, polyhedra are finite intersections of half-spaces; this corresponds to systems of linear inequalities, since one linear inequality $a \cdot x \leq b$ with $a, x \in \mathbb{R}^n, b \in \mathbb{R}$ defines the half-space $\{x : a \cdot x \leq b\}$.

geometric point of view). Thus, we will often consider a *finite market*, i.e. a market model based on a finite probability space.

Remark 23. If our model (B, S), with $S = (S^1, \dots, S^m)$, on a finite $\Omega = \{\omega_k\}_{k=1}^n$ is such that there exists $i, j \in \{1, \dots, n\}, i \neq j$ s.t. $S_1(\omega_i) = S_1(\omega_j)$, then $(B_t, S_t)(\omega_i) = (B_t, S_t)(\omega_j)$ for all $t = 0, 1$ (since B_0, B_1, S_0 are constants), i.e. our model cannot distinguish between the two points $\omega_i \neq \omega_j$. In this case, we can throw away one of those points, say ω_j, i.e. build our model on the smaller probability space $\Omega' := \{\omega_k\}_{k \in \{1,\dots,n\}, k \neq j}$, on which we consider the σ-algebra $\mathcal{A}'$ of its parts, and the probability

$$\mathbb{P}'(\omega_k) := \begin{cases} \mathbb{P}(\omega_k) & \text{if } k \neq i \\ \mathbb{P}(\omega_i) + \mathbb{P}(\omega_j) & \text{if } k = i \,. \end{cases}$$

Thus, we can and will assume w.l.o.g. that the probability space is chosen so that, given any two distinct points $\omega_i \neq \omega_j$ in it, we have $S_1(\omega_i) \neq S_1(\omega_j)$ (i.e. $S_1^k(\omega_i) \neq S_1^k(\omega_j)$ for some $k = 1, \dots, m$). Thanks to this assumption, any random variable on our probability space is[55] a function of S_1, i.e. any random variable is the payoff of a derivative.

Remark 24. We remark that in one-period models (and more generally in finite discrete time models, i.e. models with time index $\{0, 1, \dots, T\}$ for some $T \in \mathbb{N}, T \geq 1$), every complete model can[56] be built on a finite probability space. Since we are most interested in complete models, it is not much of a limitation to assume that we are working on a finite probability space (when working in finite discrete time models). As these allow also to consider incomplete models, and they allow to simplify the required probability theory (by making not reliant on measure theory), we choose to work in this setting.

If instead we consider a general probability space $(\Omega, \mathcal{A}, \mathbb{P})$, we will have to work with the set $L^0(\mathbb{P})$ of random variables (on such space), and the set of probabilities on $(\Omega, \mathcal{A})$. Here $L^0(\mathbb{P})$, and the set of measures, are infinite dimensional vector spaces with some topology. Since studying these objects is a lot more complicated from a technical point of view (in particular it requires a good knowledge of functional analysis), we will often prove theorems only under the additional (unnecessary, but strongly simplifying) assumption that the probability space is finite.

2.14 Lecture 5, How to find arbitrage

So, let us consider our usual linear model for a market (B, S), where $S = (S^1, \dots, S^m)$, on a the finite probability space $\Omega = \{\omega_i : i = 1, \dots, n\}$, and let us see how we can discover there is an arbitrage, and how to find one, relying not on intuition (which only works for simple problems) but on linear algebra. We will follow the good practices

[55]Indeed, since S_1 is (injective and thus) invertible, given any X_1 there exists f s.t. $X_1 = f(S_1)$: just define $f := X_1 \circ S_1^{-1}$, i.e. take $f := \mathrm{Im}(S_1) \to \mathbb{R}$ given by $f(s_k) := X_1(\omega_k)$ on the point $s_k := S_1(\omega_k)$.

[56]Indeed, in the market $(B, S^1, \dots, S^m)$ the replication equation has $m+1$ unknowns, and so if it always has a solution this means that the vector $(B, S^1, \dots, S^m)$ takes at most $m+1$ values, and so we can assume w.l.o.g. that it is defined on a set made of $m+1$ points. For a more formal proof see [**?**, Corollary 1.42].

of working in discounted terms, and of using vector notation for representing random variables and probabilities.

Expressed in the language of random variables, the trading strategy $(x, h) \in \mathbb{R} \times \mathbb{R}^m$ is an arbitrage if $x = 0$ and $\overline{V}_1^{0,h}$ s.t. $\overline{V}_1^{0,h} \geq 0$ a.s. and $\overline{V}_1^{0,h}$ is not a.s. $= 0$, i.e.

$$\mathbb{P}(\overline{V}_1^{0,h} \geq 0) = 1 \text{ and } \mathbb{P}(\overline{V}_1^{0,h} = 0) < 1,$$

or equivalently

$$\mathbb{P}(\overline{V}_1^{0,h} < 0) = 0 \text{ and } \mathbb{P}(\overline{V}_1^{0,h} > 0) > 0.$$

Since x always has to be 0, we often more simply say that 'h is an arbitrage'. Identifying random variables with vectors we find that, in the language of linear algebra, h is an arbitrage if the vector $w^h \in \mathbb{R}^n$, corresponding to the random variable $\overline{V}_1^{0,h} = h \cdot (\overline{S}_1 - \overline{S}_0)$, satisfies $w_i^h \geq 0$ for all i, and $w^h \neq 0$ (i.e. $w_i^h \geq 0$ for some i). Thus, if we finally really start identifying random variables as vectors and just write that $\overline{S}_1^j - \overline{S}_0^j$ is $((\overline{S}_1^j - \overline{S}_0^j)(\omega_i))_{i=1}^n$ (where, to be pedantic, instead of 'is' we should say 'is identified with'), if we denote with W the vector space of discounted payoffs replicable at cost 0, i.e.

$$W := \{\sum_{j=1}^m h^j(\overline{S}_1^j - \overline{S}_0^j) : h \in \mathbb{R}^m\} = \text{ span } \left\{(\overline{S}_1^j - \overline{S}_0^j)_{j=1}^m\right\}, \tag{8}$$

then we find that

$$\text{the set of all arbitrage payoffs is } W \cap (\mathbb{R}_+^n \setminus \{0\}); \tag{9}$$

so, there is no arbitrage iff the set $W \cap \mathbb{R}_+^n$ (which always contains the origin 0, since W is a vector space) contains only the origin (i.e. $W \cap \mathbb{R}_+^n = \{0\}$). Accidentally, we notice that the set of discounted payoffs replicable at cost x is $x + W := \{x + w : w \in W\}$, since $\overline{V}_1^{x,h} = x + \overline{V}_1^{0,h}$.

2.15 Lecture 6, The Fourier-Motzkin algorithm

As we saw above, it is important to be able to compute the set $W \cap \mathbb{R}_+^n$, and to determine if it contains just the origin. Since any vector space can be described as the set of solutions of a system of linear equalities, any vector space is a polyhedron. Since obviously the intersection of two polyhedra is a polyhedron, $W \cap \mathbb{R}_+^n$ is a polyhedron. Thus, to be able to figure out if a finite market has arbitrage, we need a method to find out if a polyhedron is empty or not.

Given a system of linear *equalities*, one way to find its solutions it to eliminate the variables one by one; the FM (Fourier-Motzkin) elimination algorithm generalises this procedure to solve a system of linear *inequalities*. We will now introduce the FM algorithm, and see how it can be used to explicitly calculate which points belong to a given polyhedron; in the next sections, we will see how to apply it to determine if a market is arbitrage-free, to find the arbitrage-free prices of a derivative, and more generally to solve LPs.

For pedagogical reasons, let us first consider the simpler case of linear equalities, already familiar to all students. Given $x \in \mathbb{R}^n, b \in \mathbb{R}^m$ and a $m \times n$ matrix A with rows $a^1, \ldots, a^m$, to solve the system $Ax = b$, we write $x = (y, z)$ with $y := (x_1, \ldots, x_{n-1}) \in \mathbb{R}^{n-1}, z := x_n \in \mathbb{R}$, so that $a^i \cdot x = c^i \cdot y + d^i z$ for some $c^i \in \mathbb{R}^{n-1}, d^i \in \mathbb{R}$, and we rewrite each equation

$$c^i \cdot y + d^i z = b_i, \quad i = 1, \ldots m$$

in the form $z = e^i \cdot y + f^i$ (if $d^i \neq 0$), or in the form $0 = e^i \cdot y + f^i$ (if $d^i = 0$), for some $e^i \in \mathbb{R}^{n-1}, f^i \in \mathbb{R}$, to find the equivalent[57] system

$$\begin{cases} z = e^i \cdot y + f^i & \text{for } i \in I_{\neq} := \{i : d^i \neq 0\} \\ 0 = e^i \cdot y + f^i & \text{for } i \in I_{=} := \{i : d^i = 0\} \end{cases} \tag{10}$$

where $I_{\neq}, I_{=}$ are two disjoint subsets whose union is $\{1, \ldots, m\}$. Notice that $I_{\neq} = \emptyset$ iff the last system, or equivalently the system $Ax = b$, does not actually involve the variable z; in this case, to find all and only the solutions (y, z) of eq. (10), we just take any y which solves the system $0 = e^i \cdot y + f^i$, $i \in I_{=}$ (which involves one fewer variable), and any $z \in \mathbb{R}$. If instead $I_{\neq} \neq \emptyset$, fix arbitrarily $i^* \in I_{\neq}$, then all and only the solutions (y, z) of eq. (10) are found by taking a solution y of the system

$$\begin{cases} e^i \cdot y + f^i = e^j \cdot y + f^j & \text{for } i, j \in I_{\neq} \\ 0 = e^i \cdot y + f^i & \text{for } i \in I_{=}, \end{cases} \tag{11}$$

which involves one fewer variable, and taking $z := e^{i^*} \cdot y + f^{i^*}$; in particular, eq. (11) has no solution iff so does $Ax = b$.

Iterating this procedure yields a sequence of linear systems $A^i x^i = b^i, i = n, n-1, \ldots, 1$, where the i^{th} system has variables $x^i := (x_1, \ldots, x_i)$, starting from $Ax = A^n x^n = b^n = b$ and deleting the last variable, one at the time, until we get to the system $A^1 x^1 = b^1$ in the only variable $x^1 = x_1 \in \mathbb{R}$, which is thus trivial to solve. Then $A^1 x^1 = b^1$ has no solution iff so does $Ax = b$, and since we can use the solutions of $A^i x^i = b^i$ to construct the solutions of $A^{i+1} x^{i+1} = b^{i+1}$, we can, by iteration, construct all the solutions of $Ax = b$ from the solutions of $A^1 x^1 = b^1$.

Example 25. To solve the system

$$\begin{cases} 2x + 3y + 0z = 6 \\ x + 3y + 0z = 1 \end{cases} \tag{12}$$

we isolate the z variable and rewrite it as

$$\begin{cases} 2x + 3y = 6 \\ x + 3y = 1 \end{cases} \tag{13}$$

[57] To systems of linear equalities/inequalities are said to be *equivalent* if they have the same set of solutions (possibly empty).

which has one fewer variable. We then isolate the y variable and rewrite the latter as

$$\begin{cases} y = & 2 - \frac{2}{3}x \\ y = & \frac{1}{3} - \frac{1}{3}x \end{cases} \tag{14}$$

This leads to the 'system' of one equation $2 - \frac{2}{3}x = \frac{1}{3} - \frac{1}{3}x$, whose unique solution is $x = 5$. Thus taking $y = 2 - \frac{2}{3}x$ (or equivalently $y = \frac{1}{3} - \frac{1}{3}x$) gives $y = -\frac{4}{3}$, and so $x = 5, y = -\frac{4}{3}$ solves eq. (13), and (x, y, z) is a solution of eq. (12) iff $x = 5, y = -\frac{4}{3}$ (z can be chosen arbitrarily).

Remark 26. Notice how we could have chosen, in eq. (13), to eliminate x instead of y. This would have been a better idea, since the resulting system

$$\begin{cases} x = & 3 - \frac{3}{2}y \\ x = & 1 - 3y \end{cases} \tag{15}$$

has fewer fractions, and thus leads to quicker calculations. Since the system here was very simple this hardly made any difference, but for systems with many variables and/or many equations choosing properly which variable to eliminate can make a real difference (when performing calculations by hand).

Let us now generalise the above procedure to the case of a system of linear inequalities, which clearly can be always be written as $Ax \geq b$. Like before, we write $x = (y, z)$ with $y := (x_1, \dots, x_{n-1}), z := x_n$, so that $a^i \cdot x = c^i \cdot y + d^i z$, and we rewrite each inequality

$$c^i \cdot y + d^i z \geq b_i, \quad i = 1, \dots m$$

in the form $z \geq e^i \cdot y + f^i$ (if $d^i > 0$), or in the form $z \leq e^i \cdot y + f^i$ (if $d^i < 0$), or in the form $0 \geq e^i \cdot y + f^i$ (if $d^i = 0$), for some $e^i \in \mathbb{R}^{n-1}, f^i \in \mathbb{R}$, to find the equivalent system

$$\begin{cases} z \geq e^i \cdot y + f^i & \text{for } i \in I_> := \{i : d^i > 0\} \\ z \leq e^i \cdot y + f^i & \text{for } i \in I_< := \{i : d^i < 0\} \\ 0 \geq e^i \cdot y + f^i & \text{for } i \in I_= := \{i : d^i = 0\} \end{cases} \tag{16}$$

where $I_<, I_>, I_=$ are disjoint sets whose union is $\{1, \dots, m\}$. Notice that $I_< \cup I_> = \emptyset$ iff the last system, or equivalently the system $Ax \geq b$, does not involve the variable z; in this case, all and only the solutions of eq. (16) are found taking y as a solution to the system $0 \geq e^i \cdot y + f^i$, $i \in I_=$, which has one fewer variable, and taking arbitrary $z \in \mathbb{R}$.

Let us now consider the case $I_< \cup I_> \neq \emptyset$. If $I_> = \emptyset$ then $I_< \neq \emptyset$, and all and only the solutions $x = (y, z)$ of eq. (16) are found by taking a solution y of the system $0 \geq d_i + f_i \cdot y$, $i \in I_=$ if $I_= \neq \emptyset$ (if $I_= = \emptyset$ then just take any $y \in \mathbb{R}^{n-1}$), and taking any $z \leq \min_{i \in I_<} d_i + f_i \cdot y$.

Analogously if $I_< = \emptyset$ then $I_> \neq \emptyset$, and we take any y solving $0 \geq d_i + f_i \cdot y$, $i \in I_=$ if $I_= \neq \emptyset$ (if $I_= = \emptyset$ then just take any $y \in \mathbb{R}^{n-1}$), and any $z \geq \max_{i \in I_>} d_i + f_i \cdot y$, and find all and only the solutions of eq. (16).

In the more interesting case where $I_>, I_<$ are both not empty, all and only solutions $x = (y, z)$ of eq. (16) are obviously obtained like this: take a solution y of the system

$$\begin{cases} e^i \cdot y + f^i \geq e^j \cdot y + f^j & \text{for } i \in I_<, j \in I_> \\ 0 \geq e^i \cdot y + f^i & \text{for } i \in I_=, \end{cases} \tag{17}$$

which involves one fewer variable, and then choose any

$$z \in [\max_{j \in I_>} e^j \cdot y + f^j, \min_{i \in I_<} e^i \cdot y + f^i]; \tag{18}$$

notice that there exists at least one such z if and only if eq. (17) has solution, and so eq. (16) has a solution iff eq. (17) does. Notice also that, if taking into account the usual conventions that $\sup \emptyset := -\infty, \inf \emptyset := \infty$ (and replacing $\max, \min$ with $\sup, \inf$), when $I_> = \emptyset$ (resp. $I_< = \emptyset$) we have that eq. (18) turns into $z \leq \min_{i \in I_<} d_i + f_i \cdot y$ (resp. $z \geq \max_{i \in I_>} d_i + f_i \cdot y$), and in particular if $I_< \cup I_> = \emptyset$ eq. (18) turns into $z \in \mathbb{R}$; so, with these conventions, no matter whether $I_<, I_>$ are empty or not, (y, z) solves eq. (16) iff y solves eq. (17) and z solves eq. (18).

Iterating this procedure shows that the FM algorithm yields a sequence of linear systems of inequalities $A^i x^i \geq b^i, i = n, n-1, \ldots, 1$, where the i^{th} system has variables $x^i := (x_1, \ldots, x_i)$, starting from $Ax = A^n x^n \geq b^n = b$ and deleting the last variable, one at the time, until we get to the system $A^1 x^1 \geq b^1$ in the only variable $x^1 = x_1 \in \mathbb{R}$, which is thus trivial to solve. Then $A^1 x^1 \geq b^1$ has no solution iff so does $Ax \geq b$, and since we can use the solutions of $A^i x^i \geq b^i$ to construct the solutions of $A^{i+1} x^{i+1} \geq b^{i+1}$, we can, by iteration, construct all the solutions of $Ax \geq b$ from the solutions of $A^1 x^1 \geq b^1$.

Example 27. Let us illustrate the use of the FM algorithm. Consider the system

$$\begin{cases} x_1 + x_2 + 2x_3 \geq 2 \\ x_1 + x_2 \geq 1 \\ x_1 - 4x_3 \geq 4 \\ 2x_1 + 3x_3 \geq 3 \\ 2x_1 - x_2 + x_3 \leq -5 \end{cases}$$

We first rewrite it as the equivalent system

$$\begin{cases} x_3 \geq 1 - \frac{1}{2}x_1 - \frac{1}{2}x_2 \\ x_3 \geq 1 - \frac{2}{3}x_1 \\ x_3 \leq -1 + \frac{1}{4}x_1 \\ x_3 \leq -5 - 2x_1 + x_2 \\ 0 \geq 1 - x_1 - x_2 \end{cases} \tag{19}$$

where notice that, when necessary, we have changed the order in which the inequalities appear, so as to have first the inequalities of the form $x_3 \geq f(x_1, x_2)$, then of the form $x_3 \leq f(x_1, x_2)$, and then of the form $0 \geq f(x_1, x_2)$.

From this last system, we build the system in only two variables

$$\begin{cases} -1+\frac{1}{4}x_1 & \geq 1-\frac{1}{2}x_1-\frac{1}{2}x_2 \\ -1+\frac{1}{4}x_1 & \geq 1-\frac{2}{3}x_1 \\ -5-2x_1+x_2 & \geq 1-\frac{1}{2}x_1-\frac{1}{2}x_2 \\ -5-2x_1+x_2 & \geq 1-\frac{2}{3}x_1 \\ 0 & \geq 1-\ \ x_1-\ \ x_2 \end{cases}$$

We now rewrite this system as

$$\begin{cases} x_2 & \geq 4-\frac{3}{2}x_1 \\ x_2 & \geq 4+x_1 \\ x_2 & \geq 6+\frac{4}{3}x_1 \\ x_2 & \geq 1-x_1 \\ 0 & \geq 2-\frac{11}{12}x_1 \end{cases} \tag{20}$$

This leads to the 'system' of one inequality $0 \geq 2-\frac{11}{12}x_1$, whose solution is any $x_1 \geq \frac{24}{11}$. Since this has solution, the original system has some solutions: let us find them all. Taking any $x_1 \geq \frac{24}{11}$, and then any

$$x_2 \geq \max\Big(4-\frac{3}{2}x_1, 4+x_1, 6+\frac{4}{3}x_1, 1-x_1\Big), \tag{21}$$

gives all solutions of eq. (20). For each such solution, taking any

$$x_3 \in [\max\Big(1-\frac{1}{2}x_1-\frac{1}{2}x_2, 1-\frac{2}{3}x_1\Big), \min\Big(-1+\frac{1}{4}x_1, -5-2x_1+x_2\Big)] \tag{22}$$

gives all solutions of eq. (19), i.e. of the system we started with. To illustrate further, if we were to specifically identify (any) one solution (x_1, x_2, x_3) of eq. (19) (e.g. to explicitly prove that it has a solution), we can choose $x_1 = 3$ (since $3 \geq \frac{24}{11}$), and then $x_2 = 10$ (since it satisfies eq. (21) with $x_1 = 3$), and then $x_3 = -1$ (since it satisfies eq. (22) $x_1 = 3, x_2 = 10$). Notice that, once chosen $x_1 = 3, x_2 = 10$, the only choice for x_3 is $x_3 = -1$; but that for other values of x_1, x_2 we could find other values of x_3 (e.g. $x_1 = 3, x_2 = 20, x_3 = -1/2$ also solves eq. (19)).

Remark 28. Unless the set of solutions of the system of linear inequalities is unique, the FM algorithm (unlike the simplex algorithm) does not provide a *convenient* description of it.

Remark 29. Since eliminating variables corresponds to computing projections, the Fourier-Motzkin algorithm can be interpreted in a geometrical way, as follows. Consider $P \subseteq \mathbb{R}^{k+m}$, and write $\mathbb{R}^{k+m}_x = \mathbb{R}^k_y \times \mathbb{R}^m_z$ to indicate that we look at $x \in \mathbb{R}^{k+m}$ as the couple $(y, z) \in \mathbb{R}^k \times \mathbb{R}^m$. Then the projection of P onto $\mathbb{R}^k_y$ is defined as being the image of P via the map $\pi_y(y, z) := y$, defined on $\mathbb{R}^k_y \times \mathbb{R}^m_z$, i.e. the set

$$\pi_y(P) := \{\pi_y(x) : x \in P\} = \{y \in \mathbb{R}^k : \exists z \in \mathbb{R}^m \text{ s.t. } (y, z) \in P\}.$$

Notice that P is empty iff $\pi_y(P)$ is empty. If $k \leq n$, let $\pi^n_k : \mathbb{R}^n \to \mathbb{R}^k$ be the projection of a vector of n coordinates onto the first k coordinates, i.e. $\pi^n_k(x) = (x_1, \ldots, x_k)$ for $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

So, what each iteration of the FM algorithm does is to compute the projection $\pi^n_{n-1}(P) = \pi_y(P)$ of a polyhedron $P \subseteq \mathbb{R}^n = \mathbb{R}^{n-1}_y \times \mathbb{R}_z$ on the[58] subspace $\mathbb{R}^{n-1}_y$, and shows that $\pi^n_{n-1}(P)$ is itself a polyhedron, by explicitly computing a representation of $\pi^n_{n-1}(P)$ as a system of linear inequalities. Notice that

$$\pi^n_k = \pi^{k+1}_k \circ \pi^{k+2}_{k+1} \circ \ldots \circ \pi^n_{n-1},$$

i.e. the elimination of the last $n-k$ variables can be achieved by eliminating the last variable $n-k$ times; so, iterating $n-k \in \{1, \ldots, n-1\}$ steps of the FM algorithm computes $\pi^n_k(P)$, and shows that it is a polyhedron.

Remark 30. The FM algorithm is one of the oldest methods for solving systems of linear inequalities. It is not very practical, because it requires a large number of steps: each time we eliminate a variable, we find a system which has potentially many more inequalities. The number of inequalities can increase fast with the number of iterations, essentially making it impossible to use this algorithm to find solutions to systems of inequalities with many variables.

For this and other reasons, normally textbooks use a different algorithm (the simplex algorithm) to solve such problems, and develop all the relative theory; this is surely the best course of action when one has a whole book to devote to the topic of LP (linear optimisation problems). For our limited purposes however, it is definitely better to use the Fourier-Motzkin algorithm: while it is much slower when dealing with LPs involving many variables and constraints, and it gives us a more limited understanding of LPs, it is far simpler, and it still allows us to solve LPs, and to prove the few theoretical results on LPs which we will need.

2.16 Lecture 7, How to find arbitrage with the FM algorithm

To figure out if a finite market has arbitrage (and, if so, to find one), we can use the FM algorithm to check whether $W \cap \mathbb{R}^n_+$ equals $\{0\}$, as we now illustrate with an example.

Example 31. Consider the trinomial model with $r = 1/9$, $S^1_0 = 5$, $S^2_0 = 10$ and

$$S^1_1 = \frac{10}{9}\begin{pmatrix} 6 \\ 6 \\ 4 \end{pmatrix}, \quad S^2_1 = \frac{10}{9}\begin{pmatrix} 12 \\ 8 \\ 8 \end{pmatrix},$$

Here the number m of stocks if 2, and the probability space has cardinality $n = 3$. Notice that in what follows S^1_0 is considered to be a random variable (with constant value 5), and is thus represented by the vector whose components are all equal to 5, and

[58] Obviously the algorithm can be easily modified to eliminate any one variable of our choice, instead of always the last variable as assumed here.

analogously for $S_0^2 = 10$. We first compute the discounted values as

$$\overline{S}_1^1 - \overline{S}_0^1 = \begin{pmatrix} 6 \\ 6 \\ 4 \end{pmatrix} - \begin{pmatrix} 5 \\ 5 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix},$$

$$\overline{S}_1^2 - \overline{S}_0^2 = \begin{pmatrix} 12 \\ 8 \\ 8 \end{pmatrix} - \begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \\ -2 \end{pmatrix},$$

and then we plug in the values for $\overline{S}$ into the eq. (6) for the discounted value of the portfolio (x, h) and get

$$\overline{V}_1^{0,h} = h^1 \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} + h^2 \begin{pmatrix} 2 \\ -2 \\ -2 \end{pmatrix}, \tag{23}$$

to find that

$$W = \{\overline{V}_1^{0,h} \Big| h \in \mathbb{R}^2\} = \text{ span } \left\{ \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \\ -2 \end{pmatrix} \right\}. \tag{24}$$

Since we are only working in dimension 3, you could now use a drawing to figure out whether $W \cap \mathbb{R}^3_+ = \{0\}$ or not; let us instead solve this with linear algebra, which works in any dimension. By definition, h is an arbitrage iff it solves the system of inequalities $\overline{V}_1^{0,h} \geq 0$ and it does *not* solve the system of equalities $\overline{V}_1^{0,h} = 0$. Keeping in mind eq. (23), we look for solutions of

$$\begin{cases} h^1 + 2h^2 \geq 0 \\ h^1 - 2h^2 \geq 0 \\ -h^1 - 2h^2 \geq 0 \end{cases} \tag{25}$$

for which not all of the above $\geq$ are satisfied with $=$. In this particular example, since the above vectors multiplying h^1 and h^2 are independent, the system $\overline{V}_1^{0,h} = 0$ has *only* the solution $h = 0$, so, any $h \neq 0$ s.t. $\overline{V}_1^{0,h} \geq 0$ is an arbitrage.

Since the system eq. (25) only has two variables, the simplest way to solve it would be to draw the intersection of the 3 half-planes represented by the 3 above inequalities. We will instead solve it using the FM algorithm (which works for any number of variables), to illustrate its use. As a first step, we isolate h^1 and get

$$\begin{cases} h^1 \geq -2h^2 \\ h^1 \geq 2h^2 \\ h^1 \leq -2h^2 \end{cases}$$

which leads to the system in the h^2 variable

$$\begin{cases} -2h^2 \geq -2h^2 \\ -2h^2 \geq 2h^2 \end{cases}$$

whose solutions are all $h^2 \leq 0$, and for each such h^2 we take any

$$h^1 \in [\max\Big(- 2h^2, 2h^2\Big), -2h^2] = \{-2h^2\}$$

and find a solution (h^1, h^2) of eq. (25). Thus, h is an arbitrage iff $h^2 < 0$, $h^1 = -2h^2$; in particular, there are arbitrages.

Week 4

2.17 Lecture 1, The no-arbitrage and the domination principles

Whenever a derivative is replicable, using the law of one price we have been able to determine at what (fair) price we ‘should’ trade it for. What should we do when a derivative is not replicable? We should use the more general assumption that there exists no arbitrage, and see where that leads us. Let us start by considering the following generalisation of the law of one price:

Principle 32 (Weak Domination Principle)**.** *If there are two possible investments, and the first one has, under all possible market outcomes*[59]*, a smaller*[60] *value then the second at time T, then this holds also at all previous times.*

This principle can be used to price derivatives; in general however it will provide us with a whole interval of ‘fair’ prices which satisfy principle 32 as follows. Suppose we can find a portfolio U which super-replicates (*resp. D which sub-replicates*) a derivative, i.e. which has a bigger (*resp. smaller*) final value *with certainty* (i.e. under all possible market outcomes). More formally, U (*resp.* D) is super-replicating (*resp. sub-replicating*) the derivative X if its value V_T^U (*resp.* V_T^U) satisfies $X_T \leq V_T^U$ (*resp.* $X_T \geq V_T^D$) a.s.[61], i.e. if the event $\{X_T > V_T^U\}$ (*resp.* $X_T < V_T^D$) has probability 0. Then principle 32 implies that:

Principle 33 (Pricing via super- and sub-replication)**.** *At any time $t \in [0, T]$, the derivative's price must be smaller (*resp. bigger*) than the value of* any *super-replicating (*resp. sub-replicating*) portfolio.*

We can then define the price bounds for the derivative as the smallest (*resp. biggest*) initial capital of a super-replicating (*resp. sub-replicating*) portfolio

$$u(X) := \inf\{V_0^U : V_T^U \geq X_T\}, \quad d(X) := \sup\{V_0^D : V_T^D \leq X_T\}; \tag{26}$$

notice that trivially a portfolio value V satisfies $V_T \geq X_T$ iff $\overline{V}_T \geq \overline{X}_T$ (and $V_T \leq X_T$ iff $\overline{V}_T \leq \overline{X}_T$), and so

$$u(X) = \inf\{V_0^U : \overline{V}_T^U \geq \overline{X}_T\}, \quad d(X) = \sup\{V_0^D : \overline{V}_T^D \leq \overline{X}_T\}, \tag{27}$$

which is a more useful expression than eq. (26), given that the formula for $\overline{V}_T$ is more convenient than that for V_T.

[59] Meaning, no matter what the value of the traded instruments turns out to be, among those values which are considered possible, i.e. which have a non-zero probability of happening.

[60] By which we always mean $\leq$, not $<$.

[61] A.s. is a common abbreviation which stands for *almost surely*; in measure theory one often uses the analogous abbreviation a.e., which means *almost everywhere*. One says that a statement hold a.s. if is holds *with probability 1*; to be precise, we should specify the probability (/the measure) in question, e.g. saying $\mathbb{P}$ a.s..

Of course, principle 32 implies that $d(X) \leq u(X)$, principle 6 is a simple corollary of principle 33, and when X is replicable at initial cost p then interval $[d(X), u(X)]$ collapses to being just a point (the point p), which is a very desirable outcome.

Principle 34 (Domination Principle)**.** *We say that* Strict Domination Principle *holds if, whenever there are two investments L, M such that*

1. *the first one has a value a.s. smaller*[62] *than the second, i.e.* $\mathbb{P}(\{V_t^L \leq V_t^M\}) = 1$
2. *the first one has a value* not *a.s. equal*[63] *to the second, i.e.* $\mathbb{P}(\{V_t^L \neq V_t^M\}) > 0$

at time $t = T$, then necessarily items 1 and 2 hold also at all previous times $t \in [0, T)$.

We will say that the Domination Principle *holds if both the law of one price 5 and the Strict Domination Principle hold.*

Of course, if the domination principle 34 holds and $V_T^L \leq V_T^M$ holds a.s., then either it holds with $=$ a.s. (and we can then apply the Law of One Price), or it does not (and we can then apply the Strict Domination Principle), so either way we can conclude $V_t^L \leq V_t^M$ a.s. for all $t \in [0, T)$, i.e. the domination principle 34 implies the weak domination principle 32. The above domination principles may or may not hold, depending on the particular market one is considering; the values of $d(X), u(X)$ depend not just on X, but also on the market that one is considering, i.e. in the portfolios which one can use to super- and sub-replicate.

Remark 35. When working in one-period models, items 1 and 2 in principle 34 are just required to hold for $t = 0$. Since V_0 is known at time 0, it is a constant, so items 1 and 2 become more simply $V_0^L < V_0^M$.

Theorem 36. *In the linear one-period market model of eq.* (5)*, the following are equivalent:*

1. *the Domination principle 34 holds.*
2. *the Strict Domination principle holds.*
3. *there exists no-arbitrage.*

Proof. The implication (1. $\Longrightarrow$ 2.) is a trivial, and so is (2. $\Longrightarrow$ 3.), since an arbitrage is an investment which, when compared[64] to the zero investment[65] violates the strict domination principle (for $t = 0$). Let us prove the implication (3. $\Longrightarrow$ 1.), by contradiction. If the domination principle fails, then either the law of one price fails, or the strict domination principle fails. If the law of one price fails, there are two portfolios

[62] i.e. the event that the first investment has value $\leq$ than the second has probability 1.

[63] i.e. the event where these two random quantities are equal does not have probability 1 (equivalently, they are different with non-zero probability).

[64] Considering the 0 investment as the first investment, an the arbitrage as the second one.

[65] i.e. to having no capital and doing nothing.

$L = (x, g), M = (y, h)$, with values V^L, V^M given as in eq. (5), s.t. $V_T^L = V_T^M$ and $x = V_0^L \neq V_0^M = y$. By remark 35, we can then assume w.l.o.g that $x > y$ (if $x < y$, just reverses the role of L and M). Thus the portfolio $M - L = (y - x, h - g)$ has initial capital $y - x < 0$ and final value $V_T^M - V_T^L = 0$. Consider now the portfolio that starts with no capital, sells g and buys h shares, i.e. the portfolio

$$N = (0, h - g) = (M - L) + (x - y, 0). \tag{28}$$

Since the final value of the portfolio $(x - y, 0)$ is $(x - y)(1 + r) > 0$, and the final wealth V_T^N of N equals

$$(V_T^M - V_T^L) + (x - y)(1 + r), \tag{29}$$

whose first term equals 0, N is an arbitrage.
Analogously, if the strict domination principle fails there are $L = (x, g), M = (y, h)$ s.t. $V_T^L \leq V_T^M$ a.s., the equality $V_T^L = V_T^M$ does not hold a.s., and yet $x \geq y$. In this case the portfolio N of eq. (28) is again an arbitrage, since the first term of eq. (29) is now ≥ 0 a.s. but not a.s. $= 0$, whereas the second term is ≥ 0. □

The above theorems shows that, while the domination principle is the most desirable property, it is actually equivalent to the no-arbitrage assumption, which is easier to check. We will say that a market satisfies NA (/the NA qcondition) if it admits no arbitrage.

The previous theorem admits the following variant, which is not nearly as useful, since what we really care about is the domination principle. To state it, we[66] define a *uniform arbitrage* as a portfolio L with zero initial capital and whose final wealth V_T^L satisfies $V_T^L \geq c\, \mathbb{P}$ a.s. (i.e. $\mathbb{P}(V_T^L < c) = 0$) for some constant $c > 0$.

Remark 37. Notice that, if the probability space Ω is finite, asking that a random variable X satisfies $X \geq c\, \mathbb{P}$ a.s. for some constant $c > 0$ is equivalent to asking that $X > 0\, \mathbb{P}$ a.s., since one can take $c := \min\{X(\omega) : \omega \in \Omega, \mathbb{P}(\{\omega\}) > 0\}$.

Theorem 38. *In the linear one-period market model of eq. (5) the Weak Domination principle 32 holds if and only if there exists no uniform arbitrage. Moreover, the Domination principle 34 implies the Weak Domination principle 32, which implies the Law of One Price, and the two opposite implications do not hold.*

Proof. If $L = (x, g)$ is a uniform arbitrage, i.e. $V_T^L \geq c$ a.s. and $x = V_0^L = 0$, then the portfolio

$$M := \Big(-\frac{c}{1 + r}, g \Big)$$

has final wealth $V_T^M = -c + V_T^L \geq 0$ a.s. and initial wealth $V_0^M = -c/(1 + r) < 0$, so the Weak Domination principle 32 fails. Conversely, if the Weak Domination principle 32 fails, there are portfolios L, M s.t. $V_T^L \leq V_T^M$ a.s. and $V_0^L > V_0^M$. The portfolio $N := M - L = (y, h)$ has final value $V_T^N = V_T^M - V_T^L \geq 0$ a.s. and initial value

[66] I made the name up, as I don't know of any source which discusses this.

$y = V_0^N = V_0^M - V_0^L < 0$. Thus the portfolio $O := (0, h) = (|y|, 0) + N$ is a uniform arbitrage, since it has initial value 0 and final value

$$V_T^O = V_T^N + |y|(1+r) \geq |y|(1+r) =: c > 0.$$

Trivially the Domination principle 34 implies the Weak Domination principle 32, and this implies the Law of One Price, since if portfolios L, M has the same final value $V_T^L = V_T^M$ then principle 32 applied to $L - M$ and to $M - L$ implies $V_0^L - V_0^M \geq 0$ and $V_0^L - V_0^M \leq 0$, and so $V_0^L - V_0^M = 0$.

Finally, let us show that the other implications do not hold. Working as in example 14 shows that a binomial model satisfies principle 32 iff $d \leq 1+r \leq u$, and so if $d = 1+r < u$ it satisfies principle 32 but not principle 34. Moreover, if $u > d > 1 + r > -1$ then principle 32 fails, yet the law of one price still holds: since the replication equation is a system of two independent equations in two unknowns, it has a unique solution. □

Remark 39. We warn the reader that the theorems in this section strongly depend on the assumptions that we are working in the linear one-period market model of eq. (5). As soon as one starts to generalise this model even slightly, and allow for 'market imperfections' (i.e., not allowing the possibility of short-selling, or considering different interest rates for borrowing and lending), the theory can change significantly, even for linear models with constraints.

2.18 Lecture 2, No-arbitrage prices

Consider the linear one-period market model (B, S) of eq. (5) in which there is no arbitrage, and an illiquid derivative X, which has a payoff X_T at maturity T. The prices at which we could then reasonably choose to trade X in this market are its 'fair' prices, defined as follows.

Definition 40. $p \in \mathbb{R}$ is a *fair* price (a.k.a. *Arbitrage-Free Price*) of X in the market (B, S) if the *enlarged* market (B, S, X), composed of the original market plus the derivative X traded has price $X_0 = p$ at time 0, is also arbitrage-free.

Often we will abbreviate Arbitrage-Free Price as AFP. Notice that pricing by the no-arbitrage principle is simply a consistency requirement (which depends on the original market (B, S)): if X is to be traded at a price p, p should be chosen in a way that does not conflict with the prices in the market (B, S), in the sense that the enlarged market (B, S, X) is still arbitrage-free.

It would be natural if replicable derivatives were the only ones with a unique fair price, i.e. for which the interval $[d(X), u(X)]$ collapses to a point. For this to be indeed the case, we need to make the following somewhat technical fact, which we will prove later on (see corollary 48) using the FM algorithm: among all the portfolios which super-replicate (*resp. sub-replicate*) a derivative, there is one with minimum (*resp. maximum*) initial value, i.e.:

Lemma 41. *The infimum and supremum in eq. (27) are attained.*

We can now use principle 33 to give an intuitive characterisation of the set of fair prices of X.

Proposition 42. *In the arbitrage-free one-period market (B, S), if a derivative X is*

1. *replicable, then its fair price X_0 is unique, it equals the initial value x of any replicating portfolio, and $u(X) = X_0 = d(X)$.*

2. *not replicable, then the set of its fair prices is the* open *interval $(d(X), u(X))$.*

Proof. In both cases we will use that, if p is a fair price for X, then by theorem 36 the Domination principle 34 holds, and so the law of one price (principle 5) holds, in the market (B, S, X) where X is sold at price $X_0 = p$.

Now, assume X is traded at price $X_0 = p$, and let us look for an arbitrage in the (B, S, X) market, i.e. for a portfolio $(0, g, h)$ with initial capital 0, g shares of S and h units of X s.t. its final wealth[67]

$$V_1^{0,g,h} = g \cdot (S_1 - (1 + r)S_0) + h(X_1 - (1 + r)X_0) \tag{30}$$

is a.s. ≥ 0 and is not a.s. $= 0$. If X can be replicated by the portfolio $M = (x, c)$ in the (B, S) market, we have that

$$X_1 = x(1 + r) + c \cdot (S_1 - (1 + r)S_0),$$

and so

$$V_1^{0,g,h} = (g + hc) \cdot (S_1 - (1 + r)S_0) + (x - p)(1 + r)h.$$

If $p = x$ then

$$V_1^{0,g,h} = (g + hc) \cdot (S_1 - (1 + r)S_0),$$

which equals the final wealth $V_1^{0,g+hc}$ of the portfolio $(0, g + hc)$ in the (B, S) market. Thus, if $(0, g, h)$ is an arbitrage in the (B, S, X) market, then $(0, g + hc)$ is an arbitrage in the (B, S) market, a contradiction. This shows that p is an arbitrage-free price of X in the (B, S) market.

Notice that x is the only fair price for X, since by the law of one price X_0 must equal the initial value of any replicating portfolio (principle 6); more precisely, if $p \neq x$ then the portfolio $(0, g, h)$ with

$$h := \frac{1}{(x - p)(1 + r)}, \quad g := -hc$$

has final wealth $V_1^{0,g,h} = 1$, so it is an arbitrage.

Since M is simultaneously super- and sub-replicating, we get that $u(X) \leq V_0^M$ and $V_0^M \leq d(X)$. Since $d(X) \leq u(X)$ always holds (by the domination principle), it follows that $d(X) = V_0^M = x = u(X)$.

[67] That this is the right formula for $V_1^{0,g,h}$ follows considering X as the $m+1$ component of S in eq. (5).

If X is not replicable and is sold at a fair price X_0, let U and D be super- and sub-replicating portfolios with extremal initial values $V_0^U = u(X)$ and $V_0^D = d(X)$, whose existence is ensured by lemma 41. The inequalities $V_T^D \leq X_T \leq V_T^U$ hold a.s., and do not hold a.s. with equality (since X is not replicable). By the domination principle $V_0^D \leq X_0 \leq V_0^U$ and (using also remark 35) $V_0^D < X_0 < V_0^U$, i.e. any fair price belongs to $(d(X), u(X))$. Let us prove that conversely any $X_0 \in (d(X), u(X))$ is a fair price for non replicable X (in the market (B, S)). Consider the portfolio[68] $N = (0, g, h)$ in the (B, S, X) market; its wealth is given by eq. (30). If N is an arbitrage then V_T^N is ≥ 0 a.s., and is not a.s. $= 0$. This implies $h \neq 0$, since otherwise $(0, g)$ would be an arbitrage in the (B, S) market, a contradiction. If $h > 0$ we get that

$$X_1 \geq X_0(1+r) - \frac{g}{h} \cdot (S_1 - S_0(1+r)) = V_1^L \tag{31}$$

holds a.s. (and $=$ does not hold a.s., though we won't use this); here L is the portfolio $(X_0, -\frac{g}{h})$ in the (B, S) market, which by eq. (31) sub-replicates X, and thus satisfies $V_0^L \leq d(X)$. Analogously if $h < 0$ we get that $V_0^L \geq u(X)$. Since $V_0^L = X_0$, we proved that if $X_0 \in (d(X), u(X))$ there cannot be any arbitrage in the (B, S, X) market, i.e. any such X_0 is a fair price for X.

□

Remark 43 (Other notions of price in incomplete models)*.* Traders need to come up with *one* price at which they should trade a derivative, not a whole interval of them. Thus, the larger the interval of fair prices for a derivative, the less useful the option pricing theory is. As the real world is of course a messy place, derivatives are never (exactly) replicable; however, as long as it is not too unreasonable to do so, it is convenient to consider models where all derivatives are replicable; these important models deserve a definition.

Definition 44. A market model (B, S) is called *complete* if any derivative X can be replicated (in such market); otherwise it is called *incomplete.*

The point is that in a complete model all derivatives have a *unique* price. The most important models we will consider (the binomial and the Black-Scholes models) are complete models. The way traders deal with incomplete models in the real world, is to use statistical considerations to pick *one* price inside the interval of arbitrage-free prices; we will not discuss how to do this in this module. Alternatively, one could consider other notions of price, which lead to a smaller interval of prices (ideally just a point). The above price bounds were derived by demanding that super- and sub-replication happen *with certainty* (i.e. a.s.), which is a very strong requirement (thus what we called fair prices would be more aptly described as not utterly unfair prices). One could instead allow for super- and sub-replication to fail 'only a little', which would lead to a smaller interval of prices. As there are many ways to interpret 'only a little', this topic is quite rich (see e.g. [**?**]), but of little relevance to how option pricing is done in the real world (as far as I know).

[68]i.e. we start with 0 initial capital, buy g units of S, and buy $h \in \mathbb{R}$ units of X at price X_0.

2.19 Lecture 3, Linear Programming, option pricing and arbitrage

Below we show how the set of arbitrage-free prices of a derivative (replicable or not) can be calculated solving a linear optimisation problem, which we now introduce.

Definition 45. A set $P \subseteq \mathbb{R}^k$ is called a *polyhedron* if it of the form

$$P := \{z \in \mathbb{R}^k : Az \geq b, \quad Cz \leq d, \quad Ez = f\},$$

where A, C, E are matrices and b, d, f are vectors. A problem of the form

$$\begin{aligned} \text{minimise} \quad & a + c \cdot z \\ \text{subject to} \quad & z \in P, \end{aligned}$$

or of the form

$$\begin{aligned} \text{maximise} \quad & a + c \cdot z \\ \text{subject to} \quad & z \in P, \end{aligned}$$

where $a \in \mathbb{R}$, c, z are vectors and P is a polyhedron, is called a *linear optimisation problem*, or a *Linear Program*, abbreviated as LP.

Remark 46. Notice that a polyhedron P can always be written in the form $P = \{z \in \mathbb{R}^k : Az \leq b\}$: indeed, clearly z satisfies $v \cdot z \leq b$ iff it satisfies $v' \cdot z \geq b'$ (with $v' = -v \in \mathbb{R}^m, b' = -b \in \mathbb{R}$), and it satisfies $v \cdot z = b$ iff it satisfies both $v \cdot z \leq b$ and $v \cdot z \geq b$. From a geometrical point of view, the set $\{z \in \mathbb{R}^k : v \cdot z \leq b\}$ is a half-space, and thus $P \subseteq \mathbb{R}^k$ is a polyhedron iff it is an intersection of finitely many half-spaces. In particular, polyhedra are always closed convex sets.

Computing the set of arbitrage-free prices boils down to solving LPs, since proposition 42 identifies the set $\mathcal{AFP}(X)$ of arbitrage-free prices of X as the open interval $(d(X), u(X))$ if $d(X) < u(X)$, and as the singleton $\{d(X)\} = \{u(X)\}$ if $d(X) = u(X)$, and $u(X), d(X)$ are solutions to some LPs, as we now show.

Indeed, since the random variable $h \cdot (\overline{S}_1 - \overline{S}_0)$ is identified with the vector whose i^{th} component is $h \cdot (\overline{S}_1 - \overline{S}_0)(\omega_i)$, calling M the matrix $M_{i,j} := (\overline{S}_1^j - \overline{S}_0^j)(\omega_i)$ we can represent the random variable $h \cdot (\overline{S}_1 - \overline{S}_0)$ as the vector Mh. If $z := (x, h) \in \mathbb{R} \times \mathbb{R}^m$ and N is the matrix whose $(k+1)^{th}$-column equals the k^{th}-column of M (for k which enumerates all columns of M), and whose first column has all elements equal to 1, then $Nz = x + Mh$, which is the vector which represents the discounted final wealth $\overline{V}_1^{x,h} = x + h \cdot (\overline{S}_1 - \overline{S}_0)$ relative to the portfolio x, h. Moreover, since the dot product between $c = (1, \mathbf{0}_m)$ and $z := (x, h) \in \mathbb{R} \times \mathbb{R}^m$ equals x (as usual $\mathbf{y}_m$ denotes the vector in $\mathbb{R}^m$ with all components equal to $y \in \mathbb{R}$), the expression eq. (27) for $u(X), d(X)$ reads

$$u(X) = \inf\{c \cdot z : z \in P_{\geq}\}, \quad d(X) = \sup\{c \cdot z : z \in P_{\leq}\}, \tag{32}$$

where the polyhedra $P_{\geq}, P_{\leq}$ are defined as

$$P_{\geq} := \{z \in \mathbb{R}^{m+1} : Nz \geq b\}, \quad P_{\leq} := \{z \in \mathbb{R}^{m+1} : Nz \leq b\},$$

where b is the vector representing the random variable $\overline{X}_1$.

Moreover, also the problem of existence of an arbitrage can be reduced to a LP. One way to see this it to notice that, since $w \in \mathbb{R}^n_+$ is the origin in $\mathbb{R}^n$ iff the dot product $w \cdot \mathbf{1}_n$ equals 0, $\nexists$ arbitrage iff the LP

$$\sup\{w \cdot \mathbf{1}_n : w \in W \cap \mathbb{R}^n_+\} \tag{33}$$

has optimal value 0. Another way is to notice that, given any polyhedron P, to determine whether P is empty or not is equivalent to determining whether the *feasibility* LP

$$\sup\{0 : z \in P\} \tag{34}$$

has optimal value (i.e. supremum) $-\infty$ or 0 (since by convention $\sup \emptyset := -\infty$).

Identifying a problem as a LP is important because LPs arises whenever one considers a linear model; in other words, LPs are ubiquitous in science, and as such they are extremely well studied. LPs have a rich and beautiful theory (called *linear programming*) behind them, and while there is no formula which produces their solution, they can be solved by hand using some simple algorithms[69]. Moreover, there are complex algorithms[70] which are able to solve LPs numerically by blazing speed, can be proved to always converge with such speed, and are capable of routinely handling LPs with hundreds of thousands of variables.

Of course, if one considers not a finite probability space but a general one, the price bounds $u(X), d(X)$ are still defined as the solutions to a linear optimisation problem, but one in infinite dimension, and so everything becomes more complicated.

2.20 Lecture 4, Solving LPs with the FM algorithm

As we have seen, if we work with a model on a finite probability space, we can price derivatives by solving the LPs (32) (whose solutions also allows us to find a super- and a sub-replicating portfolio); later we will see that one can also use an alternate set of LPs. Let us then see how can can solve LPs using the FM algorithm; by this we mean that, given a polyhedron $P \subseteq \mathbb{R}^n$ and $c \in \mathbb{R}^n$, and the LP

$$\begin{aligned} &\inf c \cdot z \\ &\text{subject to } z \in P \end{aligned} \tag{35}$$

we want to compute such infimum y^*, and find any $z^* \in P$ at which such infimum is attained (i.e. such that $c \cdot z^* = y^*$), if it exists; analogously for the problem where inf is replaced by sup. It is customary to call such z^* an *optimiser* (a.k.a. *optimal solution*, or even just *solution*) of the LP, and such y^* the *optimal value* of the LP. If an optimiser exists, the LP is called *solvable*. We recall the conventions that $\sup \emptyset := -\infty$, $\inf \emptyset := \infty$, so by definition the LP (35) has optimal value ∞ iff $P = \emptyset$.

[69] The Fourier-Motzkin algorithm, the simplex algorithm, etc.
[70] Which use the so-called interior point methods.

To solve the LP (35), we can define the polyhedron

$$R := \{(y, z) : y = c \cdot z, z \in P\} \tag{36}$$

and use the FM algorithm to compute its projection $\pi_1^{n+1}(R)$, where $\pi_1^{n+1} = \pi_y$ is defined on $\mathbb{R}^1_y \times \mathbb{R}^n_z$. Since $\pi_1^{n+1}(R) \subseteq \mathbb{R}$ is a polyhedron involving only a single variable, i.e. the set of solutions of a system of linear equations in a single variable, it is easy to handle. In particular, $\pi_1^{n+1}(R)$ is either empty (which happens iff $R = \emptyset$, i.e iff $P = \emptyset$), or a closed interval, i.e. a set of the form $(-\infty, b]$ or $[a, b]$, or $[a, \infty)$, with $a \leq b$. The infimum of $\pi_1^{n+1}(R) \subseteq \mathbb{R}$ is then easy to determine, as it equals $+\infty, -\infty, a, a$ respectively, if the polyhedron is of the form $\emptyset, (-\infty, b], [a, b], [a, \infty)$. This allows us to solve the LP (35), as follows.

Theorem 47. $\inf \pi_1^{n+1}(R)$ *equals the optimal value* y^* *of the LP* (35)*, and* z^* *solves* (35) *iff* $(y^*, z^*) \in R$.

Proof. By definition of projection, and of R, we get respectively

$$\pi_1^{n+1}(R) = \{y : \exists z \in \mathbb{R}^n \text{ s.t. } (y, z) \in R\} = \{y : \exists z \in P \text{ s.t. } y = c \cdot z\}$$

whose infimum is y^*. By definition z^* solves (35) iff $z^* \in P$ and $c \cdot z^* = y^*$, i.e. iff $(y^*, z^*) \in R$. ☐

Corollary 48. *A LP is solvable if and only if its optimal value is finite.*

Proof. Since $\sup_{z \in P} c \cdot z = -\inf_{z \in P}(-c) \cdot z$, by replacing c with $-c$ if necessary we can assume w.l.o.g. that we consider a *minimisation* LP, as in (35).

If the optimal value y^* of the LP (35) is ∞ or $-\infty$, then trivially (35) is not solvable. Conversely, if $y^* = \inf \pi_1^{n+1}(R) \in \mathbb{R}$, it means that $\pi_1^{n+1}(R)$ is of the form $[a, b], [a, \infty)$, and so $y^* = a \in \pi_1^{n+1}(R)$. We can then use the FM algorithm to compute a z^* s.t. $(y^*, z^*) \in R$ by proceeding backwards, one variable at the time (i.e., we compute z_k^* s.t. $(y^*, z_1^*, \ldots, z_k^*) \in \pi_{k+1}^{n+1}(R)$ for $k = 1$, then for $k = 2$ etc until $k = n$). Theorem 47 shows that z^* solves (35). ☐

Remark 49. Instead of the minimisation LP (35), we could analogously solve the LP

$$\begin{aligned} &\sup c \cdot z \\ &\text{subject to } z \in P \end{aligned} \tag{37}$$

with the FM algorithm, since the supremum of $\pi_1^{n+1}(R) \subseteq \mathbb{R}$ equals the optimal value y^* of the LP eq. (37), and its optimisers are the z^* s.t. $(y^*, z^*) \in R$.

Remark 50. To solve eq. (35) we could have equivalently used the polyhedron

$$R_{\geq} := \{(y, z) : y \geq c \cdot z, z \in P\}$$

instead of R, since $\inf \pi_1^{n+1}(R_{\geq})$ also equals the optimal value of eq. (35). Analogously, to solve eq. (37) we could have equivalently used the polyhedron

$$R_{\leq} := \{(y, z) : y \leq c \cdot z, z \in P\}$$

instead of R, since $\sup \pi_1^{n+1}(R_{\leq})$ also equals the optimal value of eq. (37). We cannot however use $R_{\geq}$ to solve eq. (37), nor $R_{\leq}$ to solve eq. (35).

Remark 51. When the set of optimisers of the LP (35) is not a singleton, the FM algorithm does not provide a *convenient* description of it; we only know that such set is a polyhedron, since it equals $\{z \in P : y^* = c \cdot z\}$, and it is described is some constructive, iterative fashion. For example, if the set of solutions is a vector space, it is described as the set of variables which solve a system of linear equalities, whereas a more convenient description would be in terms of determining a basis of such vector space. It is actually possible to produce an analogous and convenient description of any polyhedron, as the set spanned by the linear/conical/convex combinations of some 'basis', by applying the simplex algorithm (which we do not study, as it would take too long).

2.21 Lecture 5, How to price a derivative using the FM algorithm

Let us now see an example of how the FM algorithm can be used to compute arbitrage-free prices. Consider the following model with two stocks and a bank account with interest rate is $r = 1$. The prices at time $t = 0$ equal to $S_0^1 = 5$ and $S_0^2 = 5$. The prices of the two stocks at time $t = 1$ are given by the following vectors:

$$S_1^1 = \begin{pmatrix} 12 \\ 12 \\ 8 \\ 6 \end{pmatrix} \text{ and } S_1^2 = \begin{pmatrix} 16 \\ 8 \\ 6 \\ 4 \end{pmatrix} .$$

In this model, we want to price a call option X on S^2, with strike $K = 14$; this we do by applying the FM algorithm to compute the price bounds. To work in discounted terms, we compute:

$$\overline{S}_1^1 - \overline{S}_0^1 = \begin{pmatrix} 6 \\ 6 \\ 4 \\ 3 \end{pmatrix} - \begin{pmatrix} 5 \\ 5 \\ 5 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -2 \end{pmatrix} , \qquad \overline{S}_1^2 - \overline{S}_0^2 = \begin{pmatrix} 8 \\ 4 \\ 3 \\ 2 \end{pmatrix} - \begin{pmatrix} 5 \\ 5 \\ 5 \\ 5 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ -2 \\ -3 \end{pmatrix} , \qquad \overline{X}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} ,$$

and in particular the discounted value of the portfolio (x, h^1, h^2) at time $t = 1$ is

$$\overline{V}_1^{x,h} = \begin{pmatrix} x \\ x \\ x \\ x \end{pmatrix} + h^1 \begin{pmatrix} 1 \\ 1 \\ -1 \\ -2 \end{pmatrix} + h^2 \begin{pmatrix} 3 \\ -1 \\ -2 \\ -3 \end{pmatrix} .$$

Before we can compute the arbitrage-free prices of X, one should always check that the model is free of arbitrage; let us do that.

We write the system $\overline{V}_1^{0,h} \geq 0$, i.e.

$$\begin{cases} h^1 + 3h^2 \geq 0 \\ h^1 - h^2 \geq 0 \\ -h^1 - 2h^2 \geq 0 \\ -2h^1 - 3h^2 \geq 0 \end{cases} .$$

To minimise the number of fractions appearing, we first eliminate h^1: we write

$$\begin{cases} -3h^2 & \leq & h^1 \\ h^2 & \leq & h^1 \\ -2h^2 & \geq & h^1 \\ -\frac{3}{2}h^2 & \geq & h^1 \end{cases} \tag{38}$$

from which we get the system

$$\begin{cases} -3h^2 & \leq & -2h^2 \\ -3h^2 & \leq & -\frac{3}{2}h^2 \\ h^2 & \leq & -2h^2 \\ h^2 & \leq & -\frac{3}{2}h^2 \end{cases}$$

whose unique solution is $h^2 = 0$. To solve (38) we then take any

$$h^1 \in [\max(-3h^2, h^2), \min(-2h^2, -\frac{3}{2}h^2)] = \{0\},$$

i.e. the only solution of $\overline{V}_1^{0,h} \geq 0$ is $h = 0$, and so there is no arbitrage.

Let us now illustrate how to find the price bound $u(X)$, by solving the LP

$$u(X) = \inf \{x : (x, h) \in \mathbb{R} \times \mathbb{R}^m \text{ s.t. } \overline{V}_1^{x,h} \geq \overline{X}_1\}. \tag{39}$$

As we clarify in remark 52 below, to solve (39), we should apply the FM algorithm to the polyhedron

$$R' = \{(x, h) \in \mathbb{R} \times \mathbb{R}^m : \overline{V}_1^{x,h} \geq \overline{X}_1\}$$

eliminate the $h \in \mathbb{R}^m$ variables to compute the interval $\pi_x(R') = \pi_1^{m+1}(R')$, and then $u(X) = \inf \pi_x(R')$ is the optimal value of (39), and its optimisers are the h^* s.t. $(u(X), h^*) \in R'$. Thus, we start writing the inequalities defining R', i.e. the system

$$\begin{cases} x + \ \ h^1 + 3h^2 \geq 1 \\ x + \ \ h^1 - \ \ h^2 \geq 0 \\ x - \ \ h^1 - 2h^2 \geq 0 \\ x - 2h^1 - 3h^2 \geq 0 \end{cases}, \tag{40}$$

which is of course very similar to (but more complicated of) the system $\overline{V}_1^{0,h} \geq 0$, to which one has to first apply the FM algorithm to check for arbitrages (noticing this can speed up calculations somewhat). To minimise the number of fractions appearing, we again eliminate first h^1: we write

$$\begin{cases} -x & -3h^2 & +1 & \leq & h^1 \\ -x & +h^2 & & \leq & h^1 \\ x & -2h^2 & & \geq & h^1 \\ \frac{1}{2}x & -\frac{3}{2}h^2 & & \geq & h^1 \end{cases} \tag{41}$$

from which we get the system

$$\begin{cases} -x & -3h^2 & +1 & \leq & x & -2h^2 \\ -x & -3h^2 & +1 & \leq & \frac{1}{2}x & -\frac{3}{2}h^2 \\ -x & +h^2 & & \leq & x & -2h^2 \\ -x & +h^2 & & \leq & \frac{1}{2}x & -\frac{3}{2}h^2 \end{cases}$$

whose set of solutions is $\pi_2^3(R')$. To eliminate h^2, we rewrite this as

$$\begin{cases} h^2 & \geq & -2x & +1 \\ h^2 & \geq & -x & +\frac{2}{3} \\ h^2 & \leq & \frac{2}{3}x & \\ h^2 & \leq & \frac{3}{5}x & \end{cases} \tag{42}$$

from which we get

$$\begin{cases} \frac{2}{3}x & \geq & -2x & +1 \\ \frac{3}{5}x & \geq & -2x & +1 \\ \frac{2}{3}x & \geq & -x & +\frac{2}{3} \\ \frac{3}{5}x & \geq & -x & +\frac{2}{3} \end{cases}$$

which is equivalent to

$$\begin{cases} x & \geq & \frac{3}{8} \\ x & \geq & \frac{5}{13} \\ x & \geq & \frac{2}{5} \\ x & \geq & \frac{5}{12} \end{cases} \tag{43}$$

i.e. to $x \geq \max(\frac{3}{8}, \frac{5}{13}, \frac{2}{5}, \frac{5}{12}) = \frac{5}{12}$. Thus $\pi_1^3(R') = \pi_x(R') = [\frac{5}{12}, \infty)$, and its inf is $u(X) = \frac{5}{12}$. To find the lower price bound $d(X)$, we have to solve the LP

$$d(X) = \sup\big\{x : (x,h) \in \mathbb{R} \times \mathbb{R}^m \text{ s.t. } \overline{V}_1^{x,h} \leq \overline{X}_1\big\}. \tag{44}$$

This simply results in reversing all the inequalities (i.e. replacing $\geq$ with $\leq$, and vice versa) in all the linear systems of inequalities from (40) to (43), which leads to $x \leq \min(\frac{3}{8}, \frac{5}{13}, \frac{2}{5}, \frac{5}{12}) = \frac{3}{8}$, and so $d(X) = \frac{3}{8}$. In summary the set of arbitrage-free prices of X is the interval $(d(X), u(X)) = (\frac{3}{8}, \frac{5}{12})$.

Remark 52. While to solve (35) we started with the variable $x \in \mathbb{R}^n$ and defined the additional variable $y = c \cdot x$ to build the polyhedron R as in eq. (36), in the LP (39) the variables are $(x,h) \in \mathbb{R} \times \mathbb{R}^m$, and we need to minimise x, so there is no point in defining the additional variable y, since $c = (1, \mathbf{0}_m)$ and so $y = c \cdot (x,h) = x$. In other words, if we were to solve (39) by blindly following the procedure outlined when considering the generic LP (35), we would apply the FM algorithm to the set

$$R := \{(y,x,h) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^m : y = x, (x,h) \in R'\},$$

we would compute $y^* := \inf \pi_y(R)$, and work backward to find (x^*, h^*) s.t. $(y^*, x^*, h^*) \in R'$, and use theorem 47 to conclude that y^* is the optimal value and (x^*, h^*) the optimal solution of (39). However, we can more simply solve (35) by applying the FM algorithm to R' to compute $a^* := \inf \pi_x(R')$, and then work backward to find b^* s.t. $(a^*, b^*) \in R'$: indeed, we claim that $a^* = y^*$ and (a^*, b^*) is the optimal solution of (39). To prove this, notice that, since the projection on the (y, x) plane of R is

$$\pi_{(y,x)}(R) = \{(y, x) : y = x, x \in \pi_1^3(R')\} = \{(x, x) : x \in \pi_1^3(R')\},$$

from which it follows that

$$\pi_y(R) = \pi_y(\pi_{(y,x)}(R)) = \pi_1^3(R') = \pi_x(R'). \tag{45}$$

Thus, if $a^* := \inf \pi_x(R')$ and $(a^*, b^*) \in R'$, then eq. (45) gives $a^* = y^*$, and so the fact that

$$(y, x, h) \in R \iff y = x, \quad (x, h) \in R',$$

which holds simply by definition of R, shows that $(y^*, y^*, b^*) \in R$, i.e. $(y^*, b^*) = (a^*, b^*)$ is the optimal solution of (39) (by theorem 47).

2.22 Lecture 6, How to find the optimisers of a LP using the FM algorithm

Let us illustrate how, using the FM algorithm, once determined the optimal value $y^* \in \mathbb{R}$ of (35), we can compute its optimisers. By definition of projection, the polyhedron

$$\{z \in \mathbb{R} : (y^*, z) \in \pi_2^{n+1}(R)\} = \pi_z(\pi_2^{n+1}(R))$$

is not empty, and so it is a closed interval, and so we can easily find a point $z = z_1^*$ in it: for example, if $\pi_2^{n+1}(R)$ was given by eq. (16), from which we are to eliminate the variable z to get to the polyhedron $\pi_1^{n+1}(R)$, which is the set of solution of eq. (17), then z_1^* would be any z given by eq. (18). By the same reasoning,

$$\{z \in \mathbb{R} : (y^*, z_1^*, z) \in \pi_3^{n+1}(R)\} = \pi_z(\pi_3^{n+1}(R))$$

is a closed interval, and we can find a point z_2^* in it, etc. Iterating like this, we eventually find a point $(y^*, z_1^*, \dots, z_n^*) \in \pi_{n+1}^{n+1}(R) = R$, and now Theorem 47 shows that z^* solves (35). Notice that at every step, we could *choose* which point z_i^* to consider, within the polyhedron $\pi_z(\pi_{i+1}^{n+1}(R))$; as we have shown, all possible such choices leads to an optimiser z^*. Conversely, any optimiser z^* is obtained in this way, since it satisfies

$$z_i^* \in \{z \in \mathbb{R} : (y^*, z_1^*, \dots, z_{i-1}^*, z) \in \pi_{i+1}^{n+1}(R)\}$$

if $i = n$ (because $(y^*, z^*) \in R = \pi_{n+1}^{n+1}(R)$), and so also if $i < n$ (by definition of projection).

To clarify the above procedure, we let us consider again the example in section 2.21, and illustrate how to find the solutions (x, h) of the LP (39), i.e. the portfolios which super-replicate X starting with initial capital $x = \frac{5}{12}$. Using (42) and substituting $x = \frac{5}{12}$ we find

$$h^2 \in \Big[\max(-2x+1, -x+\frac{2}{3}), \min(\frac{2}{3}x, \frac{3}{5}x)\Big] = \Big[\max(\frac{1}{6}, \frac{1}{4}), \min(\frac{5}{18}, \frac{1}{4})\Big] = \Big[\frac{1}{4}, \frac{1}{4}\Big] = \{\frac{1}{4}\}$$

and finally using (41) we find

$$h^1 \in \Big[\max(-x - 3h^2 + 1, -x + h^2), \min(x - 2h^2, \frac{1}{2}x - \frac{3}{2}h^2)\Big]$$

and substituting $x = \frac{5}{12}$ and $h^2 = \frac{1}{4}$ we get

$$h^1 \in \Big[\max(-\frac{5}{12} - \frac{3}{4} + 1, -\frac{5}{12} + \frac{1}{4}), \min(\frac{5}{12} - \frac{1}{2}, \frac{5}{24} - \frac{3}{8})\Big]$$

and so

$$h^1 \in \Big[\max(-\frac{1}{6}, -\frac{1}{6}), \min(-\frac{1}{12}, -\frac{1}{6})\Big] = \{-\frac{1}{6}\}$$

In summary, the LP (39) has the unique solution (i.e. optimiser)

$$x = \frac{5}{12}, \quad h^1 = -\frac{1}{6}, \quad h^2 = \frac{1}{4},$$

and optimal value $x = \frac{5}{12}$.

Week 5

2.23 Lecture 1, Foreign currency

The FX (Foreign eXchange) market is huge and very liquid, since companies which sell their products/services in several markets receive income in multiple currencies, and then want to convert it to their domestic currency. This exposes them to *FX risk*, i.e. the risk that comes from the fact that exchange rates change unpredictably over time. To avoid such risk, companies commonly trade options based on exchange rates.

Example 53. Consider that Spotify, which sells subscriptions in € to its EU-based customers, also sells subscriptions to its UK-based customers, who pay in £. Spotify, which is headquartered in Sweden, may want a predictable revenue measured in €, yet it does not want to charge its UK costumers a monthly fee which changes monthly, so as to track the exchange rate $E := E_{£}^{€}$ between £ and € (defined as the cost of one € in £). Instead, to get a predictable revenue in €, Spotify could buy a forward contract on the exchange rate, which fixes in advance the rate E at which it will exchange a certain amount M of £ to € at time T. It could then choose M to be the profits generated by its UK operations between now and maturity T.

Given that the value of the exchange rate at the future time $t > 0$ is unpredictable, we describe it with a random variable E_t. Given that such value E_t changes over time, the exchange rate E should then be modelled with a (strictly positive) stochastic process $E = (E_t)_{t \geq 0}$.

Recall that we chose to describe the money market using the simple idealisation of 'the bond' (or 'the bank account'). Analogously, if we can trade in two currencies, we should consider a domestic bond B^d, whose value is given in £, and a foreign bond B^f, valued in the foreign currency. These two bonds would normally have different interest rates, though both of their values at time 0 can assumed to be 1 (by normalisation). So, in one period models, we take

$$B_0^d = 1, \quad B_0^f = 1, \quad B_0^d = 1 + r^d, \quad B_0^f = 1 + r^f,$$

where $r^d, r^f > -1$ are the domestic and foreign interest rates. On top of the domestic and foreign bond, the market under consideration could also include domestic assets, valued in the domestic currency, and foreign assets, valued in the foreign currency. When considering a market with assets valued in multiple currencies, one should always measure all values in the same currency, so as not to 'compare apples with oranges'. So, consider an investment (e.g. bond, stock, etc.) in Europe, whose price (in €) is S; its price in £ is then SE, where $E := E_{£}^{€}$ is the cost of one € in £. Conversely, if U is the value in £ of an investment, its value UE' in €, where the exchange rate $E' := E_{€}^{£}$ between € and £, which is defined as the cost of one £ in €, obviously satisfies $E_{€}^{£} = 1/E_{£}^{€}$. Notice that one cannot invest directly in exchange rates, so the processes E, E' are not to be considered as part of the market.

Example 54. Consider as possible investments 'the UK and EU bonds', with values B^d, B^f, a British stock S^d, and a European stock S^f. Obviously the values of B^d, S^d are given in £, while those of of B^f, S^f are given in €. Since the price in £ of the EU bond and stock are EB^f and ES^f, where $E := E^{€}_{£}$, a British investor would model this market with the processes (B^d, S^d, EB^f, ES^f), whereas a Europeans one would model the same market as $(B^d/E, S^d/E, B^f, S^f)$. Moreover, the British investor would normally use B^d as numeraire, whereas the European one would use B^f.

To clarify further, compare these two simple possible investments. At time 0 I could deposit €1 in the foreign bank, in which case at time 1 I'd have €$(1+r^f)$, which I could convert to £$E_1(1+r^f)$. Alternatively, at time 0 I could convert the €1 in £ to get £E_0, and then deposit that in the domestic bank account would yield £$E_0(1+r^d)$ at time 1.

Finally, we remark that clearly the qualitative properties of the market do not depend on the currency in which one does the accounting. For example, in the setting of example 54, (B^d, S^d, EB^f, ES^f) is arbitrage free/complete iff $(B^d/E, S^d/E, B^f, S^f)$ is such.

2.24 Lecture 2, Formulas for the binomial model

Let is consider for the moment the one-period binomial model (B, S). W.l.o.g. we assume that $S_0 d = S_1(T) < S_1(H) = S_0 u$, as it is convention to do; we do *not* yet assume that the model is free of arbitrage, i.e. $d < 1+r < u$ (see example 14). To price a derivative, we can to try to replicate it, i.e. we look for x, h such that

$$\overline{V}_1^{x,h}(H) = \overline{X}_1(H), \quad \overline{V}_1^{x,h}(T) = \overline{X}_1(T). \tag{46}$$

Since $\overline{V}_1^{x,h} = x + h(\overline{S}_1 - \overline{S}_0)$, this system of two (independent, since $u \neq d$) equations is in the two unknowns x, h, and so for every derivative the system has a unique solution, i.e. the binomial model is complete. In particular, if $d \geq 1+r$ or $1+r \geq u$, we have built an example of a complete model with arbitrage; and if you assign to each derivative as price the initial value of a replicating portfolio, we have a model which satisfies the law of one price but does not satisfy the[71] domination principle.

Let us find the formula that gives h (*resp.* x) as a function of $\overline{X}_1(H), \overline{X}_1(T)$ by taking linear combinations of the two eq. (46) so as to obtain an equation only in h (*resp.* x), making easy to solve. We first consider

$$h(\overline{S}_1(H) - \overline{S}_1(T)) = \overline{V}_1^{x,h}(H) - \overline{V}_1^{x,h}(T) = \overline{X}_1(H) - \overline{X}_1(T),$$

from which we get the *delta-hedging formula*

$$h = \frac{X_1(H) - X_1(T)}{S_1(H) - S_1(T)}. \tag{47}$$

[71] If moreover the inequality $d \geq 1+r$ (or $1+r \geq u$) holds strictly, the model does not even satisfy the weak domination principle

To find x we instead fix $\tilde{p}$ and consider

$$\tilde{p}\overline{V}_1^{x,h}(H) + (1-\tilde{p})\overline{V}_1^{x,h}(T) = \tilde{p}\overline{X}_1(H) + (1-\tilde{p})\overline{X}_1(T),$$

or equivalently

$$x + h\Big((\tilde{p}\overline{S}_1(H) + (1-\tilde{p})\overline{S}_1(T)) - \overline{S}_0\Big) = \tilde{p}\overline{X}_1(H) + (1-\tilde{p})\overline{X}_1(T),$$

and call $C_{\tilde{p}}$ the quantity multiplying h on the LHS of the above equation. Then we choose $\tilde{p}$ such that $C_{\tilde{p}} = 0$, and so we get

$$x = \tilde{p}\overline{X}_1(H) + (1-\tilde{p})\overline{X}_1(T). \tag{48}$$

This equation provides the value of x given $\tilde{p}$; to find $\tilde{p}$ we solve

$$0 = C_{\tilde{p}} := (\tilde{p}\overline{S}_1(H) + (1-\tilde{p})\overline{S}_1(T)) - \overline{S}_0. \tag{49}$$

Written in terms of the interest rate r, and of the up and down factors $u := S_1(H)/S_0$ and $d := S_1(T)/S_0$, the solution of $0 = C_{\tilde{p}}$ is

$$\tilde{p} := \frac{(1+r)-d}{u-d}. \tag{50}$$

Notice that the value of $\tilde{p}$ does not depend on X_1, and that $d < 1+r < u \iff \tilde{p} \in (0,1)$. Thus, if $d < 1+r < u$ and we define

$$\tilde{q} := 1-\tilde{p} = \frac{u-(1+r)}{u-d}, \quad \mathbb{Q}(H) := \tilde{p}, \quad \mathbb{Q}(T) := \tilde{q}, \tag{51}$$

we have build a probability $\mathbb{Q}$ on $\{H,T\}$ which satisfies $\mathbb{Q}(\{\omega\}) > 0$ for all $\omega \in \Omega = \{H,T\}$, and satisfies $\overline{S}_0 = \mathbb{E}^{\mathbb{Q}}\overline{S}_1$, since this is just eq. (49) in a different notation. Conversely, if there exists a probability $\mathbb{Q}$ on $\{H,T\}$ which satisfies $\mathbb{Q}(\{\omega\}) > 0$ for all $\omega \in \{H,T\}$, and $\overline{S}_0 = \mathbb{E}^{\mathbb{Q}}\overline{S}_1$, then $\tilde{p}$, defined in eq. (50), satisfies $\tilde{p} = \mathbb{Q}(H) \in (0,1)$, and so $d < 1+r < u$. In summary, the binomial model has no arbitrage iff there exists a probability $\mathbb{Q}$ s.t. $\mathbb{Q}(\{\omega\}) > 0$ for $\omega \in \{H,T\}$, and $\overline{S}_0 = \mathbb{E}^{\mathbb{Q}}\overline{S}_1$, and in this case the price $X_0 = \overline{X}_0$ of any derivative is given by the *Risk-Neutral Pricing Formula*

$$\overline{X}_0 = \mathbb{E}^{\mathbb{Q}}\overline{X}_1, \tag{52}$$

which is just eq. (48) in a different notation

Example 55. In the binomial model with $r = 1, S_0 = 6, S_1(H) = 18, S_1(T) = 2$, let us price and hedge the derivative with payoff X_1 at time 1 given by $X_1(H) = 4, X_1(T) = 16$, in the following two ways:

1. by using the risk neutral pricing formula and the delta hedging formula

2. by computing by hand a replication strategy and its cost

First notice that the up and down factors are $u := S_1(H)/S_0 = 3, d := S_1(T)/S_0 = 1/3$, and so $d < 1+r < u$, which means that this binomial model is arbitrage free (a fact that you should always check, before trying to price something). Let us now apply each of the two above mentioned methods to solve the problem, so we can then compare them.

1. The risk neutral probability $\mathbb{Q}$ is given by

$$\mathbb{Q}(H) = \frac{2-\frac{1}{3}}{3-\frac{1}{3}} = \frac{6-1}{9-1} = \frac{5}{8}, \quad \mathbb{Q}(T) = 1 - \mathbb{Q}(H) = \frac{3}{8},$$

so the risk neutral pricing formula gives the price

$$X_0 = \frac{1}{1+r}\mathbb{E}[X_1] = \frac{1}{2}\left(\frac{5}{8}\cdot 4 + \frac{3}{8}\cdot 16\right) = \frac{17}{4},$$

and the delta hedging formula states that to hedge X_1 one needs to buy

$$h = \frac{X_1(H) - X_1(T)}{S_1(H) - S_1(T)} = \frac{4-16}{18-2} = -\frac{3}{4}$$

stocks (i.e. shortsell 3/4 stocks), starting with initial capital $X_0 = \frac{17}{4}$ and putting the remaining cash $X_0 - hS_0 = \frac{17}{4} + \frac{3}{4}\cdot 6 = \frac{35}{4}$ in the bank.

2. Setting $V_1 = kB_1 + hS_1$ equal to X_1 gives the following system

$$\begin{cases} 2k + 18h = 4 \\ 2k + 2h = 16 \end{cases}$$

whose solution is $k = \frac{35}{4}, h = -\frac{3}{4}$, which corresponds to starting with initial cash $kB_0 + hS_0 = \frac{35}{4}\cdot 1 - \frac{3}{4}\cdot 6 = \frac{17}{4}$.

Remark 56. Both of the above approaches can be generalised to work in general models, not just the binomial one. With this in mind, let us compare them.

The replication equation approach is much more intuitive than the EMM one. Solving the replication equation not only allows to find out whether a derivative is replicable, it also explicitly provides a replicating strategy when there is one. The EMM method instead tells us whether a derivative is replicable or not, and which are its prices, finding actually finding the replicating strategy: for that one has to additionally apply the delta-hedging formula.

When working on a finite probability space, solving the replication equation leads to somewhat less computations than the EMM method if one has to price only *one replicable* derivative; when there are two of more derivatives to be priced, the EMM method is quicker. Indeed, while the first replication method requires to solve a system of equations *for each replicable derivative* (and two systems of inequalities for each non-replicable derivative, as explained below), the EMM method requires to solve a system only once (to find the EMM), and then to evaluate an integral for each replicable derivative (and then also to find the range of a function for each non-replicable derivative). Analogously, notice that when an option is not replicable, the replication equation cannot be solved, so to find the AFP one has to solve two systems of inequalities to find instead the smallest super-replication and largest sub-replication prices, which is at least twice as laborious as solving the replication equation; so, also in this case it is quicker to use the EMM approach.

There are at least two very considerable advantages of using EMMs. One is that, when working in continuous-time models, the replication approach is hard to implement, whereas the EMM approach is as straightforward as in discrete time; for example, when working in the Black and Scholes model (which is complete), one has a nice formula[72] for the (unique) EMM $\mathbb{Q}$ (analogous to eqs. (50) and (51)), and to price a derivative with payoff X_T one simply needs to compute $\mathbb{E}^{\mathbb{Q}}\overline{X}_T$. Then, using a delta-hedging formula somewhat analogous to eq. (47), one can use the formula for the price of the derivative to find the hedging strategy.

Another advantage is that, when working with models with market imperfections, principle 34 and proposition 42 do not hold, and yet one can generalise the EMM approach to hold also in this settings.

Remark 57. While it is easy to remember that to find $\mathbb{Q}$ we just need to solve $\overline{S}_0 = \mathbb{E}^{\mathbb{Q}}\overline{S}_1$, it may be a good idea to remember by heart the formulas for $\tilde{p}$ and $\tilde{q}$; to do so, it helps to think as follows. Assume that there is no-arbitrage; then $\tilde{p}$ and $\tilde{q}$ are strictly positive, add up to one, and depend on the parameters $d, 1+r, u$ (not on S_0) of the model, which satisfy $d < 1+r < u$. It is then reasonable that they should be given by how long the subintervals $[d, 1+r]$ and $[1+r, u]$ are relative to the whole interval $[d, u]$.

2.25 Lecture 3, The Fundamental Theorem of Asset Pricing

Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, if $\mathbb{Q}$ is another probability on $(\Omega, \mathcal{A})$, $\mathbb{Q}$ is said to be *absolutely continuous* with respect to $\mathbb{P}$ (in symbols $\mathbb{Q} \ll \mathbb{P}$), if $\mathbb{P}(A) = 0 \Longrightarrow \mathbb{Q}(A) = 0$ for any $A \in \mathcal{A}$, i.e. if any null set of $\mathbb{P}$ is a null set of $\mathbb{Q}$. The probabilities $\mathbb{P}, \mathbb{Q}$ are said to be *equivalent* (in symbols $\mathbb{Q} \sim \mathbb{P}$) if $\mathbb{Q} \ll \mathbb{P}$ and $\mathbb{P} \ll \mathbb{Q}$, i.e. if they have the same null sets. Notice that, if Ω is finite (or countable) and every singleton $\{\omega\}$, $\omega \in \Omega$, is measurable, then $\mathbb{Q} \ll \mathbb{P}$ holds iff $\mathbb{P}(\{\omega\}) = 0$ implies $\mathbb{Q}(\{\omega\}) = 0$; in particular, if Ω is finite and $\mathbb{P}(\{\omega\}) > 0$ for all $\omega \in \Omega$ then any probability $\mathbb{Q}$ satisfies $\mathbb{Q} \ll \mathbb{P}$, and $\mathbb{Q}$ satisfies $\mathbb{Q} \sim \mathbb{P}$ iff $\mathbb{Q}(\{\omega\}) > 0$ for all $\omega \in \Omega$.

Thus, in section 2.24 we have proved that the (one-period) binomial model is arbitrage-free if and only if there exists a probability $\mathbb{Q}$ on $\{H, T\}$ which is equivalent to $\mathbb{P}$ and such that $\overline{S}_0 = \mathbb{E}^{\mathbb{Q}}\overline{S}_1$ holds. This suggests the following definition and theorem. In all that follows $(\Omega, \mathcal{A}, \mathbb{P})$ is an arbitrary probability space (not assumed finite), on which are defined some processes $B, S = (S^1, \dots, S^m)$ s.t. $B > 0$, and as usual $\overline{W} := W/B$ denotes discounting.

We recall that, if Y is a positive random variable then its expectation $\mathbb{E}Y \in [0, \infty]$

[72]More precisely, the formula is for the Radon-Nikodym density $d\mathbb{Q}/d\mathbb{P}$ and, more importantly, this allows to obtain the law of the underlying under $\mathbb{Q}$, which is one what one needs to compute $\mathbb{E}^{\mathbb{Q}}\overline{X}_T$.

is always defined[73], though it could take the value ∞. If X is a random variable s.t. $\mathbb{E}|X| < \infty$ (i.e. the expectation of its absolute value is finite), then we say that X is $\mathbb{P}$-integrable, and we denote the (vector) space of all $\mathbb{P}$-integrable random variables with $L^1(\mathbb{P})$. This space matters since the expectation $\mathbb{E}X \in \mathbb{R}$ is defined for all $X \in L^1(\mathbb{P})$. If we want to clarify that the expectation is with respect to the probability $\mathbb{P}$ we can write $\mathbb{E}^{\mathbb{P}}$ instead of $\mathbb{E}$; thus, $\mathbb{E}^{\mathbb{Q}}$ denotes the expectation with respect to the probability $\mathbb{Q}$.

Definition 58. A probability $\mathbb{Q}$ on $(\Omega, \mathcal{A})$ is a $\overline{S}$-*Martingale Measure* (a.k.a. a Martingale Measure for $\overline{S}$/for (B, S), or a *risk-neutral measure*), if $\mathbb{Q} \ll \mathbb{P}$ and

$$\overline{S}_1^j \in L^1(\mathbb{Q}), \qquad \overline{S}_0^j = \mathbb{E}^{\mathbb{Q}}(\overline{S}_1^j), \quad \forall j = 1, \dots, m. \tag{53}$$

A martingale measure equivalent to $\mathbb{P}$ is called an *Equivalent Martingale Measure.*

We will normally just use the abbreviations MM and EMM for the above probabilities. We will denote with

$$\mathcal{M}(\overline{S}) := \{\mathbb{Q} \sim \mathbb{P} : eq.\ (53) \text{ holds }\}, \quad \mathbb{M}(\overline{S}) := \{\mathbb{Q} \ll \mathbb{P} : eq.\ (53) \text{ holds }\}, \tag{54}$$

the families of all EMM/MM for $\overline{S}$, often abbreviated as $\mathcal{M}, \mathbb{M}$.

Remark 59 (MM in finite Ω). Any random variable defined on a finite Ω is integrable with respect to any probability; thus, if Ω is finite, the assumption $\overline{S}_1^j \in L^1(\mathbb{Q})$ is automatically satisfied. Moreover, because of our assumption that $\mathbb{P}(\{\omega\}) > 0$ for all $\omega \in \Omega$, $\mathbb{Q} \ll \mathbb{P}$ is satisfied by any probability $\mathbb{Q}$, and $\mathbb{Q} \sim \mathbb{P}$ holds iff $\mathbb{Q}(\{\omega\}) > 0$ for all $\omega \in \Omega$.

Remark 60 (Why $\mathbb{Q} \sim \mathbb{P}$). It is obvious that, if for any reason we are to ever consider another probability $\mathbb{Q}$ on $(\Omega, \mathcal{A})$, it is essential that $\mathbb{Q} \ll \mathbb{P}$. Indeed, since all that we know about the model are statements[74] that have probability 1 under $\mathbb{P}$, i.e. we don't know anything about sets that have probability 0 under $\mathbb{P}$, it is then essentially that these sets have probability 0 also under $\mathbb{Q}$ (i.e. are irrelevant also under $\mathbb{Q}$). Analogously the property $\mathbb{P} \ll \mathbb{Q}$ allows us to deduce that if some facts hold under $\mathbb{Q}$, than they hold under $\mathbb{P}$. So, it is not surprising that the property $\mathbb{Q} \sim \mathbb{P}$ (i.e. $\mathbb{Q} \ll \mathbb{P}$ *and* $\mathbb{P} \ll \mathbb{Q}$) of being equivalent to $\mathbb{P}$ is important. Sometimes however it is useful to consider the (more general) absolutely continuous martingale measures, i.e. MM s.t. $\mathbb{Q} \ll \mathbb{P}$.

Remark 61. Note that definition of arbitrage (and thus of arbitrage-free prices) does not involve the full knowledge of the probability measure $\mathbb{P}$ (i.e. it does not depend on the *exact* values which $\mathbb{P}(A)$ takes), as it only depend on the family of null[75] sets of $\mathbb{P}$. One could then hope to switch from $\mathbb{P}$ to another measure $\mathbb{Q} \sim \mathbb{P}$, under which all AFPs will be the same, and such that using $\mathbb{Q}$ simplifies the calculations. This is exactly what we have done, since using $\mathbb{Q}$ it suffices to use eq. (52) to price a derivative. A risk-neutral

[73]For example one can define it as $\sup_{n \in \mathbb{N}} \mathbb{E}[\min(Y, n)]$, once $\mathbb{E}$ has been defined for bounded random variables.

[74]Statements like 'this strategy super-replicates that payoff', or like 'this strategy is an arbitrage'.

[75]A null set is a set of probability 0.

measure is then just a mathematical object, which has no relation with the real world, and it is to be seen as simply an artefact which conveniently allows us to simplify proofs and calculations.

Remark 62 (Risk-Neutral measures). Since any investor is risk-averse, i.e. is only willing to take more risks in exchange for higher returns, in any realistic model the average return[76] of any stock should[77] be higher than the return of the bond, or equivalently $\mathbb{E}^{\mathbb{P}}(\overline{S}_1^j) > S_0^j$ for all j; here the average is taken using the *physical measure* $\mathbb{P}$, i.e. the probability that an investor considers appropriate to describe the probability of events in the real world. This shows that the physical measure does not satisfy eq. (53); if the investor was instead indifferent to risk, and all (s)he cared about was average returns, than eq. (53) would hold. This is why a $\mathbb{Q}$ s.t. eq. (53) is called a risk-neutral measure, and why eq. (52) holds: if both the stock and the bond have the same average return r, than so does every portfolio of stocks and bonds. Though in reality the physical measure is subjective (i.e. depends on the investor), in our models we instead just take one fixed $\mathbb{P}$ for all investors; this would clearly not be a reasonable assumption if different investors had different information about the market.

We have seen how to characterise the validity of the domination principle via the lack of arbitrage. Generalising the result which we obtained in Section 2.24 for the binomial model, the absence of arbitrage can also be characterised using EMMs, as follows.

Theorem 63 (1st Fundamental Theorem of Asset Pricing). *Consider the linear one-period market model (B, S) of eq. (5). This model is free of arbitrage $\iff \mathcal{M}(\overline{S}) \neq \emptyset$.*

In the very special and simple case of the binomial model, we have proved the implication $\implies$ of the above crucial theorem (normally abbreviated as FTAP), which is the difficult one. We will not present here the proof of this implication for general Ω, as it is very technical (we refer the interested reader to [**?**, Theorem 1.7]). However, we will present one possible proof in the case where Ω is finite, using the following classic theorem, which has an extremely intuitive statement.

Theorem 64 (Separation Theorem). *If $C, K \subseteq \mathbb{R}^n$ are convex, C is closed and K is compact then there exists $z \in \mathbb{R}^n, a, b \in \mathbb{R}$ s.t.*

$$x \cdot z \leq a < b \leq y \cdot z, \quad \forall x \in C, y \in K.$$

In particular, if C is a vector space then $x \cdot z = 0$ for all $x \in C$.

Proof. Here only the idea: just take $z := c^* - k^*$, where c^*, k^* are minimisers of the optimisation problem $\min_{c \in C, k \in K} ||c - k||$, where $|| \cdot ||$ is the usual (Euclidean) norm on $\mathbb{R}^n$. If C is a vector space then $x \cdot z = 0$ for all $x \in C$, since otherwise by linearity for every $b \in \mathbb{R}$ there would exists $x \in C$ s.t. $x \cdot z > b$, a contradiction. □

[76] The return of a portfolio M is the random variable R defined by the equation $V_0^M(1 + R) = V_1^M$.

[77] In our models we do not assume that $\mathbb{E}(\overline{S}_1^j) > S_0^j$ holds, since this would not help us in any way.

The above separation theorem can be generalised to many[78] infinite dimensional vector spaces (e.g. to Banach spaces), in which case it is known as the Hahn-Banach theorem, which is probably the most important theorem in analysis.

Proof of the 1st FTAP. ($\Longleftarrow$) Let us assume by contradiction that h is an arbitrage, so $\overline{V}_1^{0,h} \geq 0$ $\mathbb{P}$ a.s. and so $\mathbb{Q}$ a.s., and $\{\overline{V}_1^{0,h} > 0\}$ is not a null set under $\mathbb{P}$ and thus also under $\mathbb{Q}$; it follows that $\mathbb{E}^{\mathbb{Q}}(\overline{V}_1^{0,h}) > 0$. This contradicts the fact that, since $\mathbb{Q}$ is a MM for (B, S),

$$\mathbb{E}^{\mathbb{Q}}(\overline{V}_1^{0,h}) = h \cdot (\mathbb{E}^{\mathbb{Q}}(\overline{S}_1) - \overline{S}_0) = 0.$$

($\Longrightarrow$) We only sketch the proof of this implication, and only for finite $\Omega = \{\omega_i\}_{i=1}^n$. By eq. (9) the set W of discounted wealths attainable at cost 0, and the set $\mathbb{R}_+^n \setminus \{0\}$, are disjoint, so a fortiori $W \cap \Delta_n = \emptyset$, where $\Delta_n := \{q \in \mathbb{R}_+^n : \sum_{i=1}^n q_i = 1\}$ is compact and convex. From theorem 64 it follows that $\exists z$ s.t. $0 = w \cdot z < b \leq y \cdot z$ for all $w \in W, y \in \Delta_n$. Since the elements $(e_i)_i$ of the canonical basis of $\mathbb{R}^n$ belong to Δ_n, this implies $0 < e_i \cdot z = z_i$ for all i, so $c := \sum_{i=1}^n z_i > 0$. Thus $q := z/c \in \mathcal{M}$, since $0 = w \cdot q$ means $0 = \mathbb{E}^Q \overline{X}_1$ for all $\overline{X}_1 \in W$. $\square$

Remark 65. Theorem 63 gives us one way to prove whether there exists an arbitrage: one needs to solve the system of linear equations

$$\mathbf{1}_n \cdot q = 1, Mq = 0, \quad q \in \mathbb{R}^n$$

and then impose the condition $q_i > 0$ for all i to compute the set of $q \in \mathcal{M}$.

Here is yet another way. Recall that

$$\mathcal{M} = \{q \in \mathbb{R}^n : q_i > 0 \quad \forall i, \, \mathbf{1}_n \cdot q = 1, Mq = 0\},$$

and define

$$\mathcal{D} := \{z \in \mathbb{R}^n : z \geq \mathbf{1}_n, Mz = 0\}. \tag{55}$$

Since the function $z(q) := \frac{1}{\min_i q_i} q$ maps $\mathcal{M}$ to $\mathcal{D}$, and the function $q(z) := \frac{1}{\sum_i z_i} z$ maps $\mathcal{D}$ to $\mathcal{M}$, we see that $\mathcal{M} \neq \emptyset$ iff $\mathcal{D} \neq \emptyset$, and so by theorem 63

$$\exists \text{ arbitrage} \iff \mathcal{D} = \emptyset. \tag{56}$$

This is useful since $\mathcal{D}$ (unlike $\mathcal{M}$) is a polyhedron, so to determine if it is empty we just need to solve a LP (the feasibility problem); so, this second method is in a way better than the first, when working on a finite probability space. Of course, in such a setting, one may as well attack the problem directly, as done in section 2.16; this is not harder, and has the added advantage of explicitly computing an arbitrage when there is one. However, if working with an infinite probability space (which is essential in the continuous-time setting) theorem 63 proves to be an invaluable result, as it is *the* way in which one proves that a model admits no arbitrage.

[78] In holds for locally-convex topological vector spaces.

Analogously to theorem 63, one can use martingale measures to characterise the existence of a uniform arbitrage (i.e. the failure of the weak domination principle) as follows.

Theorem 66. *Consider the linear one-period market model* (B, S) *of eq.* (5)*, and assume that* Ω *is finite. In this model there exists no uniform arbitrage* $\iff \mathbb{M}(\overline{S}) \neq \emptyset$*.*

Now we present only the proof of the simple implication $\Longleftarrow$. If Ω is finite, the opposite implication can be easily proved using the theory of linear programming (in this setting also theorem 63 can be proved using linear programming, though its proof is harder). The correct way to generalise theorem 66 so as to have a statement for general (infinite) Ω is considered in [**?**, Theorem 3.9].

Proof. ($\Longleftarrow$) Let us assume by contradiction that h is an uniform arbitrage, so $\overline{V}_1^{0,h} \geq c > 0$ $\mathbb{P}$ a.s.. and so[79] $\mathbb{Q}$ a.s.; it follows that $\mathbb{E}^{\mathbb{Q}}(\overline{V}_1^{0,h}) \geq c > 0$. This contradicts the fact that, since $\mathbb{Q}$ is a MM for (B, S),

$$\mathbb{E}^{\mathbb{Q}}(\overline{V}_1^{0,h}) = h \cdot (\mathbb{E}^{\mathbb{Q}}(\overline{S}_1) - \overline{S}_0) = 0.$$

□

Remark 67. There many possible notions of arbitrage which have been studied in the literature, and ultimately one should pick one based not just on the financial interpretation but also choosing a notion that leads to a good mathematical theory (e.g. a nice characterisation as in theorem 63). Thus, normally people choose either the notion of *no free lunch with vanishing risk*, or the notion of *no unbounded profit with bounded risk*, which have been characterised with a FTAP-like theorems in [**?**] and [**?**]. These differences however are too technical to be of interest to us, and what we care about is theorem 63, not theorem 66. We have mainly mentioned the notion of uniform arbitrage to draw attention to the following details: the difference between the domination principle and the weak domination principle, and correspondingly between EMM and MM, and between the open and closed intervals $(d(X), u(X))$ and $[d(X), u(X)]$ of prices, appearing in Section 2.26, for which no arbitrage (either standard or uniform) exists.

2.26 Lecture 4, The Risk Neutral Pricing Formula

We assume throughout that we are working in the linear one-period market model (B, S) of eq. (5). The following immediate corollary of theorem 63 is most important.

Corollary 68 (Risk-Neutral Pricing Formula)**.** *The set of arbitrage-free prices for an illiquid derivative with payoff* X_1 *in a one-period arbitrage-free market model* (B, S) *is*

$$\mathcal{AFP}(X_1) = \{\mathbb{E}^{\mathbb{Q}}[\overline{X}_1] : \mathbb{Q} \in \mathcal{M}(\overline{S}) \text{ and } \mathbb{E}^{\mathbb{Q}}[|\overline{X}_1|] < \infty\}.$$

[79] Since $\mathbb{Q} \ll \mathbb{P}$, the set $\{\overline{V}_1^{0,h} < c\}$ is also a null set under $\mathbb{Q}$.

Proof. By theorem 63 the set $\mathcal{M}(\overline{S})$ is non-empty, and $X_0 = \overline{X}_0$ is an AFP for X iff the market (B, S, X), where X is traded at price X_0 at time 0, has no arbitrage, i.e. iff the set $\mathcal{M}(\overline{S}, \overline{X}) \subseteq \mathcal{M}(\overline{S})$ is also non-empty. The thesis then follows from the fact that by definition $\mathbb{Q} \in \mathcal{M}(\overline{S})$ belongs to $\mathcal{M}(\overline{S}, \overline{X})$ iff $\mathbb{E}^{\mathbb{Q}}[|\overline{X}_1|] < \infty$ and $\overline{X}_0 = \mathbb{E}^{\mathbb{Q}}[\overline{X}_1]$. □

The same proof, applied using theorem 66 instead of theorem 63, leads to following alternative version of the RNPF.

Corollary 69 (Risk-Neutral Pricing Formula)**.** *If (B, S) is a one-period market model with no uniform arbitrage, based on a finite*[80] *probability space, the set of prices for an illiquid derivative with payoff X_1 for which the extended market (B, S, X) has no uniform arbitrage equals*

$$\{\mathbb{E}^{\mathbb{Q}}[\overline{X}_1] : \mathbb{Q} \in \mathbb{M}(\overline{S}) \text{ and } \mathbb{E}^{\mathbb{Q}}[|\overline{X}_1|] < \infty\}. \tag{57}$$

Remark 70. It follows from corollary 68 and proposition 42 that the price bounds for X are given by

$$u(X) = \sup\{\mathbb{E}^{\mathbb{Q}}(\overline{X}_1) : \mathbb{Q} \in \mathcal{M}(\overline{S}), \mathbb{E}^{\mathbb{Q}}[|\overline{X}_1|] < \infty\}, \tag{58}$$

$$d(X) = \inf\{\mathbb{E}^{\mathbb{Q}}(\overline{X}_1) : \mathbb{Q} \in \mathcal{M}(\overline{S}), \mathbb{E}^{\mathbb{Q}}[|\overline{X}_1|] < \infty\}, \tag{59}$$

The sup and inf in eqs. (58) and (59) are attained if and only if X_1 is replicable. Indeed, by corollary 73 they are trivially attained if X_1 is replicable. If instead X_1 is not replicable, by proposition 42 the set of its AFPs is the *open* interval $(d(X), u(X))$, and so by corollary 68 $\{\mathbb{E}^{\mathbb{Q}}[\overline{X}_1] : \mathbb{Q} \in \mathcal{M}(\overline{S})\} = (d(X), u(X))$, which does not have a minimum nor a maximum.

Remark 71. Assume that $\Omega = \{\omega_i\}_{i=1}^n$ is finite, so $\mathbb{M}$ is (identified with) the polyhedron

$$\mathbb{M} = \{q \in \mathbb{R}^n : q \geq 0, \mathbf{1}_n \cdot q = 1, Mq = 0\},$$

where $\mathbf{1}_n^T$ denotes the transpose of the vector with n components all equal to 1, and as usual $M_{j,i} := (\overline{S}_1^j - \overline{S}_0^j)(\omega_i)$. In this case, $\mathbb{M}$ is compact: any polyhedron is obviously closed, and $\mathbb{M}$ is bounded since $\mathbb{M} \subseteq [0, 1]^n$.

If $\mathbb{Q} \in \mathbb{M}$, then $\mathbb{Q} \in \mathcal{M}$ iff the inequalities $\mathbb{Q}(\omega_i) > 0$ are satisfies strictly for all i, and so $\mathcal{M}$ is (identified with) the set

$$\mathcal{M} = \{q \in \mathbb{R}^n : q_i > 0 \quad \forall i, \mathbf{1}_n \cdot q = 1, Mq = 0\}.$$

If $\mathbb{R} \in \mathcal{M}$ then $\mathbb{Q}_t := t\mathbb{R} + (1 - t)\mathbb{Q} \in \mathcal{M}$ for any $t \in (0, 1]$, and since[81] $\mathbb{Q}_t \to \mathbb{Q}$ as $t \downarrow 0$ we find that $\mathcal{M}$, *if not empty*[82], is dense in $\mathbb{M}$. Thus, if $\mathcal{M} \neq \emptyset$, $\mathbb{M}$ is[83] the closure of $\mathcal{M}$.

[80] If Ω is not finite, one should assume not the absence of uniform arbitrage, but rather the absence of weak instantaneous profit, since this is equivalent to $\mathbb{M}$ being non-empty, see [**?**, Theorem 3.9].

[81] With convergence in the sense of vectors in $\mathbb{R}^n$, i.e. $\mathbb{Q}_t(\omega_i) \to \mathbb{Q}(\omega_i)$.

[82] It can of course happen that there exists MMs but there do not exists any EMMs: for example this happens in the binomial model with $d = 1 + r < u$, as it can easily be proved directly (is also follows from theorems 38 and 66).

[83] Since $\mathcal{M}$ is dense in $\mathbb{M}$, and $\mathbb{M}$ is closed.

Remark 72. Assume that Ω is finite. Then, any random variable on Ω is integrable with respect to any probability, so the sup and inf in eq. (58) and eq. (59) are over the set of all $\mathbb{Q} \in \mathcal{M}(\overline{S})$. If we assume that (B, S) is arbitrage-free, then $\mathbb{M}$ is the closure of $\mathcal{M}$ (by theorem 63 and remark 71), and so we can equivalently[84] replace EMM with MM in eq. (58) and eq. (59). This is convenient for two reasons.

First, in this case the sup and inf are then attained: since $\mathbb{Q} \mapsto \mathbb{E}^{\mathbb{Q}}(\overline{X}_1)$ is (linear and thus) continuous, its inf and sup on the compact set $\mathbb{M}$ are attained. In particular, the set which appears in eq. (57) is the *closed* interval $[d(X), u(X)]$. The inf and sup are attained by MM which are not equivalent to $\mathbb{P}$, by remark 70.

Second, $\mathbb{M}$ being a polyhedron, eqs. (58) and (59) state that we can obtain the price bounds as a solution to the LPs

$$u(X) = \max\{b \cdot q : q \in \mathbb{M}\}, \quad d(X) = \min\{b \cdot q : q \in \mathbb{M}\}, \tag{60}$$

where b is the vector $\overline{X}_1$ (i.e. $b_i = \overline{X}_1(\omega_i)$ for $i = 1, \dots, n$). This gives us an alternative to using the LP eq. (32) to calculate the price bounds.

We can easily compute if X_1 is replicable without even having to find the replicating strategy (by solving the replication equation), as follows.

Corollary 73 (Characterisation of replicable derivatives)**.** *X_1 is replicable $\iff \mathbb{E}^{\mathbb{Q}}[\overline{X}_1]$ is constant across all $\mathbb{Q} \in \mathcal{M}(\overline{S})$ s.t. $\mathbb{E}^{\mathbb{Q}}[|\overline{X}_1|] < \infty$.*

Proof. This trivially follows from proposition 42 and corollary 68. □

The following simple characterisation of complete models is often called the 2nd Fundamental Theorem of Asset Pricing.

Corollary 74. *Let (B, S) be a market model free of arbitrage. Then (B, S) is complete $\iff$ the EMM is unique (i.e. $\mathcal{M}(\overline{S})$ is a singleton).*

Proof. ($\Longleftarrow$) If the EMM $\mathbb{Q}$ is unique then by corollary 73 any derivative is replicable, i.e. the market is complete. ($\Longrightarrow$) Given an arbitrary $A \in \mathcal{A}$, since the derivative with payoff $X_1 = 1_A$ is replicable, by corollary 68 we have that, given any $\mathbb{Q}^1, \mathbb{Q}^2 \in \mathcal{M}(\overline{S})$,

$$\mathbb{Q}^1(A)/(1+r) = \mathbb{E}^{\mathbb{Q}^1}[\overline{X}_1] = \mathbb{E}^{\mathbb{Q}^2}[\overline{X}_1] = \mathbb{Q}^2(A)/(1+r).$$

This shows $\mathbb{Q}^1(A) = \mathbb{Q}^2(A)$ for all $A \in \mathcal{A}$, i.e. $\mathbb{Q}^1 = \mathbb{Q}^2$, i.e. $\mathcal{M}(\overline{S})$ is a singleton. □

[84] Since the function $\mathbb{Q} \mapsto \mathbb{E}^{\mathbb{Q}}(\overline{X}_1)$ is (linear and thus) continuous, its inf and sup over $\mathcal{M}$ yield the same value as over its closure $\mathbb{M}$.

2.27 Lecture 5, Dividends

To simplify matters, we have so far ignored a complication which presents itself when pricing options whose underlying is a stock or a bond. Some[85] companies, instead of investing all of its profits (in its own operations, in the stock market, etc.), choose to hand out some of them in cash to its share-holders; such payments are called *dividends*, and are of the two main reasons why people buy stocks (the other being the hope to realise capital gains, i.e. to be able to resell the shares at a higher price at a later time). The stock-holder can then reinvest such cash in the same stock by buying more shares, or deposit it in the bank account, or use it in other ways. Although the size of such payments (called *dividend rate*, or simply just *dividend*), and times of payment (called *ex-dividend dates*), can depend on several quantities (time, stock price, market conditions, etc.), in general the dividends get paid every 3 months, always in the same amount.

Analogously, a bond can specify that its buyer is entitled not just to receive a large payment (called *principal*, or *face value*, or *par value*, or simply *par*) at maturity, but also smaller payments (called *coupons*) at intermediate times. A bond paying no coupons is called a *zero-coupon bond*. Bonds can have variable coupons, but many bonds pay fixed coupons, every 6 months.

Of course, paying out dividends decreases the value of the shares by the same amount. Thus, if we denote with S the price of one share just *after* the dividend D is paid (called *ex-dividend*, or *post-dividend*) and V the price of one share just *before* the dividend D is paid (called *cum-dividend*), then clearly $V = S + D$, $S \geq 0, D \geq 0$.

Let us now mention some complications created by considering dividends, in relation to arbitrage, numeraire, short-selling, and option pricing.

To define arbitrage, one has to keep in mind that the value of a portfolio holding one share depends also of the value of the associated dividends, and of how they have been reinvested. In particular, when working in a one-period model, it is the cum-dividend price that one has to use in order to determine whether there exists an arbitrage. For example, consider the one-period binomial model with $r = 3$, one asset with initial price $S_0 = 2$, and ex-dividend price S_1 which can take values $4, 8$, and assume that the dividend is the constant $D = D_1 = 2$ (paid at time 1; no dividend is paid at time 0). Then this model is arbitrage free, even if the two values u, d taken by S_1/S_0 do not satisfy $d < 1 + r < u$, because what matters is that the two values u', d' taken by V_1/V_0 do indeed satisfy $d' < 1 + r < u'$ (since $d' = (4 + 2)/2 = 3, 1 + r = 4, u' = (8 + 2)/2 = 5$).

Analogously, one should use V, not S, as a numeraire, because it is V which represents the value of a portfolio, not S.

We then point out that after you short-sell a share, when you have to return the share to the broker you borrowed it from, you owe him/her also compensation for the dividends issued by the stock in the meantime.

We now show with an example how the pricing of options is affected by dividends,

[85] Normally only well established companies, with predictable profits, pay dividends, whereas companies in the early stages of development prefer to invest any profits back into their business to grow faster.

by calculating the forward price F of a stock paying dividends in a one-period model. Clearly F depends on the dividend D, because while the forward contract pays $S_1 - F$, which does not depend on D, the stock does pay dividends. Notice that if we buy one share at cost S_0, while borrowing $\frac{F+D}{1+r}$ from the bank, we replicate the payment of $S_1 - F$ of the forward contract, since the final wealth corresponding to this trading strategy is

$$(S_1 + D) - (1 + r)\left(\frac{F + D}{1 + r}\right) = S_1 - F.$$

Since the initial cost of such replicating strategy is $S_0 - \frac{F+D}{1+r}$, setting this to 0 shows that the forward price of S is $F = S_0(1 + r) - D = 2 \cdot (1 + 3) - 2 = 6$, which does indeed depend on D as expected.

3 Multi-Period models

Week 6

3.1 Lecture 1, Measurability

We will consider in this section multi-period market models. To consider these, we need to consider the notion of a stochastic processes; this is a family $X = (X_t)_{t\in I}$ (i.e. a function $I \ni t \mapsto X_t$) of random (real) numbers or vectors, all defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and indexed by some set I. In the case of real numbers, X_t is a random variable; in the case of vectors in $\mathbb{R}^n$, $X_t = (X_t^1, \ldots, X_t^n)$ is a random vector, i.e. a finite family of random variables.

When we write that two processes X, Y satisfy $X \le Y$ we mean that $X_t \le Y_t$ holds for all $t \in I$ a.s., i.e.

$$\mathbb{P}(\{\omega : X_t(\omega) \le Y_t(\omega)\, \forall t \in I\}) = 1; \tag{61}$$

notice that the condition $X_t \le Y_t$ holds a.s. for all $t \in I$, i.e.

$$\mathbb{P}(\{\omega : X_t(\omega) \le Y_t(\omega)\}) = 1, \quad \forall t \in I,$$

is equivalent[86] to the condition in eq. (61) whenever I is countable (like when $I = \mathbb{T}$), and thus also[87] when $I = [0, \infty)$ *if X, Y are continuous* (but not in general). Analogous definitions are used for the relations $=, \ge, <, >$ between processes.

Normally one considers $I \subseteq [0, \infty]$ as representing time: now we will consider the case $I = \mathbb{T} := \{0, 1, \ldots, T\}$ (in continuous time one takes $I = [0, T]$), with $T \in \mathbb{N}$ (one could also consider $T = \infty$, but we won't as it would lead to complications). Thus our multi-period market model will be described by the (vector-valued) stochastic process $(B, S^1, \ldots, S^m)$ indexed by $\mathbb{T}$, where $B > 0$ represents the value of the bond and each S^j the value of the j^{th} underlying.

One new difficulty is that we need to introduce a way to express, using mathematics, the intuitive concept that a trading strategy $(K, H) = (K_t, H_t)_{t\in\mathbb{T}}$ is *non-anticipative*, i.e. for each t, the random quantities K_t, H_t can depend only on the past, i.e. on information known (before or) at time t. We could do this by declaring what is the set of random variables $\mathcal{S}_t$ which are known at time t, and then asking that (K_t, H_t) is a (non-random) function of those random variables, so that all the randomness in (K_t, H_t) comes from $\mathcal{S}_t$, i.e. once the value of all random variables in $\mathcal{S}_t$ is known, the value of (K_t, H_t) is known, is not anymore random. So for example if we assume that $\mathcal{S}_t = \{B_u, S_u\}_{u\in\mathbb{T}, u\le t}$, we could ask that for every t there exists a function f (which can depend on t) s.t.

$$(K_t, H_t) = f(t, B_0, S_0, \ldots, B_t, S_t).$$

[86]Since the union of *countably many* null sets (i.e. sets of probability 0) is a null set.

[87]Because the set $\mathbb{Q} \cap [0, \infty)$ is countable, and by the continuity assumption

$$\{\omega : X_t(\omega) \le Y_t(\omega)\, \forall t \in [0, \infty)\} = \{\omega : X_t(\omega) \le Y_t(\omega)\, \forall t \in \mathbb{Q} \cap [0, \infty)\}$$

The above definition has the problem that, for some function f, the quantity

$$f(t, B_0, S_0, \ldots, B_t, S_t)$$

is not a random variable. This problem is solved by restricting f to being a Borel-measurable function; let us briefly review what this means, by also introducing the concept of generated σ-algebra, which we will also need later on. The following two remarks are highly technical, and the reader should not worry about the proofs of any statements made therein.

Remark 75 (Generated sigma algebra). It follows trivially from the definition that the intersection $\mathcal{G} := \cap_{j \in J} \mathcal{G}_j$ of σ-algebras $\mathcal{G}_j, j \in J$, where J is *any* set, is a σ-algebra. Thus, given an arbitrary collection $\mathcal{C}$ of subsets of Ω, one can define the σ-algebra $\sigma(\mathcal{C})$ generated by $\mathcal{C}$ as being the intersection of all σ-algebras which contain $\mathcal{C}$ (there exists at least one such σ-algebra: the family of parts of Ω, i.e. of all subsets of Ω). It is obviously the smallest σ-algebra containing $\mathcal{C}$, meaning that if $\mathcal{C}' \supseteq \mathcal{C}$ is a σ-algebra then $\mathcal{C}' \supseteq \sigma(\mathcal{C})$.

Given $f : A \to B$ and a σ-algebra $\mathcal{B}$ on B, the family $\{f^{-1}(B) : B \in \mathcal{B}\}$ is a σ-algebra, denoted by $\sigma(f)$. If $\mathcal{A}$ is a σ-algebra on A, then f is said to be $\mathcal{A}/\mathcal{B}$-measurable if $\sigma(f) \subseteq \mathcal{A}$. More generally, given $f_j : A \to B, j \in J$, the family of sets

$$\cup_{j \in J} \{f_j^{-1}(B) : B \in \mathcal{B}\},$$

is not necessarily a σ-algebra, but one can consider the σ-algebra it generates, which is denoted with $\sigma((f_j)_{j \in J})$; it is obviously the smallest σ-algebra $\mathcal{A}$ on A for which every $f_j, j \in J$ is $\mathcal{A}$-$\mathcal{B}$ measurable.

Remark 76 (The Borel σ-algebra). If $\mathcal{C}$ is the family of all open sets of $\mathbb{R}^n$, the generated σ-algebra $\sigma(\mathcal{C})$ defined in remark 75 is called the Borel σ-algebra, and denoted with $\mathcal{B}(\mathbb{R}^n)$. One could equivalently take many other choices of $\mathcal{C}$ to generate the Borel sets, e.g. the family of all closed sets, or of all rectangles[88], or of all balls[89]; one could even additionally impose that each set is bounded, that rectangles/balls are open (or closed), and that each interval has rational endpoints and each ball has rational centre and radius. Recall that a function $X : \Omega \to \mathbb{R}^n$ defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is called a random vector (or random variable, if $n = 1$) if it is $\mathcal{A}/\mathcal{B}(\mathbb{R}^n)$-measurable. One can show that

$$\sigma(X) := \{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R}^n)\}$$

equals

$$\sigma(X_1, \ldots, X_n) := \sigma(\cup_{i=1}^n \{X_i^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}),$$

and so X is a random vector iff every component X_i of $X = (X_i)_{i=1}^n$ is a random variable. A set $B \subseteq \mathbb{R}^n$ is called *Borel*-measurable if $B \in \mathcal{B}(\mathbb{R}^n)$, and a function $f : \mathbb{R}^n \to \mathbb{R}^k$ is called *Borel*-measurable if it is $\mathcal{B}(\mathbb{R}^n)/\mathcal{B}(\mathbb{R}^k)$-measurable. In this class,

[88] A rectangle $R \subseteq \mathbb{R}^n$ is a set of the form $R = I_1 \times \ldots \times I_n$, where each $I_i, i = 1, \ldots, n$ is an interval.

[89] A ball $B \subseteq \mathbb{R}^n$ is a set of the form $\{x \in \mathbb{R}^n : ||x - c|| < r\}$ (open ball) or of the form $\{x \in \mathbb{R}^n : ||x - c|| \leq r\}$ (closed ball), where $c \in \mathbb{R}^n$ is the center, $r \geq 0$ the radius, and $|| \cdot ||$ the Euclidean norm.

and throughout much of mathematics, $\mathbb{R}^n$ is always endowed with its Borel σ-algebra, so one can simply state that $B \subseteq \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}^k$ are measurable to mean Borel-measurable; alternatively, one can also simply say Borel, instead of Borel-measurable. Because the composition of measurable functions is measurable, $f(X)$ is a random vector and $g \circ f$ is Borel if X is a random vector and f, g are Borel (where $g : \mathbb{R}^k \to \mathbb{R}^m$). Because the Borel σ-algebra is 'very big', it is very hard to construct a non-Borel set, and a non-Borel function, so essentially you will never encounter them (unless you search for counter-examples in a book); for example, any continuous function is Borel, and the (pointwise) limit of Borel functions is Borel. However, unfortunately one really has to make the assumption of Borel-measurability, because one cannot define a reasonable notion of volume (Lebesgue measure) on the σ-algebra of all subsets of $\mathbb{R}^n$, e.g. because of the Banach-Tarski paradox (look it up, it is insanely awesome!).

Remark 77. In probability and measure theory, one always identifies sets with $\{0,1\}$-valued functions, using the bijection

$$A \mapsto 1_A =: X, \text{ with inverse } X \mapsto \{\omega : X(\omega) = 1\} =: A,$$

where the function

$$1_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \in \Omega \setminus A \end{cases}$$

is called the *indicator* of A.

It is possible to reformulate the above condition in an equivalent way which is less intuitive, but more convenient to work with, as we show in the next lemma.

Lemma 78 (Doob-Dynkin). *Suppose X and Y are random vectors with n and k components, defined on the measurable space $(\Omega, \mathcal{A})$, then the following are equivalent:*

1. *There exists a Borel function $f : \mathbb{R}^k \to \mathbb{R}^n$ such that $X = f(Y)$*
2. *X is $\sigma(Y)$-measurable, i.e. $\sigma(X) \subseteq \sigma(Y)$*

Proof. The implication (1) $\Longrightarrow$ (2) is trivial. While we won't give a full proof of the opposite implication (which can be found in any book on measure theory), we sketch here the idea. Given a σ-algebra $\mathcal{G}$, obviously $A \in \mathcal{G}$ iff 1_A is $\mathcal{G}$-measurable. As we have characterised the sets in $\sigma(Y)$ as being those of the form $\{Y \in B\} := Y^{-1}(B)$, for some $B \in \mathcal{B}(\mathbb{R}^k)$, a set A is in $\sigma(Y)$ iff $1_A = 1_B(Y)$ for some $B \in \mathcal{B}(\mathbb{R}^k)$. Thus, by remark 77, we proved the thesis when X is a $\{0,1\}$-valued random variable. By taking linear combinations of such functions, we see that the lemma holds whenever X takes finitely many values. By taking limits of such functions, the results holds for any X. □

To verify the condition in lemma 78 becomes particularly easy when Y only takes countably many values $\{y_n\}_{n \in \mathbb{N}}$, using the next simple lemma.

Lemma 79. *Given random vectors X, Y on the same measurable space $(\Omega, \mathcal{A})$, if Y only takes countably many values $\{y_n\}_{n \in \mathbb{N}}$ then the t.f.a.e.*

1. *There exists a function $f : \mathbb{R}^n \to \mathbb{R}^k$ such that $X = f(Y)$*

2. *X takes a constant value x_n on each set of the form $\{Y = y_n\}, n \in \mathbb{N}$.*

In this case one has to choose f s.t. $f(y_n) = x_n$ for all $n \in \mathbb{N}$, and can choose $f = 0$ on $\mathbb{R}^k \setminus \{y_n\}_{n\in\mathbb{N}}$, in which case f is Borel-measurable.

Proof. If $X = f(Y)$ then $X = f(y_n)$ on $\{Y = y_n\}$ for each $n \in \mathbb{N}$. Conversely, if $X = x_n$ on $\{Y = y_n\}$ for each $n \in \mathbb{N}$, then defining f as

$$f(y) := \begin{cases} x_n & \text{if } y = y_n, n \in \mathbb{N} \\ 0 & \text{if } y \in \mathbb{R}^k \setminus \{y_n\}_{n\in\mathbb{N}} \end{cases}$$

we obtain a function f s.t. $X = f(Y)$. Such f is trivially Borel, since any set of the form $f^{-1}(B)$ is of the form $\cup_{n\in K}\{y_n\}$ or of the form $(\cup_{n\in K}\{y_n\}) \cup (\mathbb{R}^k \setminus \{y_n\}_{n\in\mathbb{N}})$ (for some $K \subseteq \mathbb{N}$), and every singleton is a Borel set. □

Because of lemma 78, in probability theory, information is modelled using σ-algebras, and we can interpret $\sigma(Y)$ as encoding all the information contained in Y. Since we assume that information accumulates over time (i.e. that no information is ever lost), we should consider a σ-algebra $\mathcal{F}_t$ (which represents the information known at time t) which is increasing in t. We can now formalise the above discussion with some definitions.

Definition 80. A family $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ of σ-algebras is called a *filtration* if $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s \leq t, s, t \in \mathbb{T}$.

Definition 81. The stochastic process $X = (X_t)_{t\in\mathbb{T}}$ is said to be *adapted* to the filtration $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ (or *$\mathcal{F}$-adapted*) if X_t is $\mathcal{F}_t$-measurable for all $t \in \mathbb{T}$. The *natural filtration* generated by a process X is the smallest filtration $\mathcal{F}^X$ to which X is adapted, i.e. $\mathcal{F}_t^X = \sigma((X_u)_{u\leq t, u\in\mathbb{T}})$.

So, given a market model (B, S) on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, we will always endow such space with a filtration $\mathcal{F}$ to which (B, S) is adapted: unless stated otherwise, we will choose the natural filtration of (B, S). It could make sense to include a larger filtration, as information other than past asset prices could be of value in determining future asset prices (e.g. information about weather, political events etc.). The quadruple $(\Omega, \mathcal{A}, \mathcal{F}, \mathbb{P})$ is then called a *filtered probability space*; in fact, since the filtration is so basic to the definition of a stochastic process, the more modern term for a filtered probability space is a *stochastic basis*.

When trading, we will only consider strategies (K, H) which are adapted to $\mathcal{F}$ (since (K, H) should be non-anticipative), and a process X represents the stream of payoffs of a derivative iff it is adapted to the filtration generated by the underlying and the bond. Moreover, the interest rate $R_t > -1$ which one receives for investing money in the bond between time t and time $t + 1$ should be known at time t, i.e. R must be adapted; thus $B_t = B_0(1 + R_0) \cdot \ldots \cdot (1 + R_{t-1})$ should be known at time $t - 1$, i.e. B must be *$\mathcal{F}$-predictable*, i.e. B_t is $\mathcal{F}_{t-1}$-measurable for all $t \in \mathbb{T}, t > 0$. One can interpret the bond being predictable as being less risky: its value is known one time step in advance. To be precise H_t and R_t are not defined at time $t = T$.

We warn the reader that some authors call H_{t+1} what here we call H_t (i.e. the number of shares between time t and time $t + 1$), and correspondingly they demand that H be predictable; this is just a matter of taste.

We will often assume that $R_t = r$ is constant in time (and so $B_t = B_0(1 + r)^t$) and deterministic (i.e. not random); while this is of course an unrealistic and extreme over-simplification, it can be a convenient one.

3.2 Lecture 2, Self-financing portfolios

If at time t we own $H_t^j \in \mathbb{R}$ units of the j^{th} asset with price S_t^j, and K_t units of the bond with price B_t, at time t our portfolio's value will be $K_tB_t + H_t \cdot S_t$, and at time $t + 1$ it will be $K_tB_{t+1} + H_t \cdot S_{t+1}$. At time $t + 1$ we will then re-adjust our portfolio, and decide to own K_{t+1} bonds and H_{t+1} shares between time $t + 1$ and $t + 2$. The values of our portfolio just before and just after this re-adjustment at time $t+1$, are $K_tB_{t+1}+H_t \cdot S_{t+1}$ and $K_{t+1}B_{t+1} + H_{t+1} \cdot S_{t+1}$, and of course they should be equal, since we are assuming that while trading we are not consuming our wealth nor investing new wealth, we are just re-investing, shuffling it around between the bonds and the various underlying, the wealth that we have thus far accumulated by trading when starting from some initial wealth x. Thus, we require the our portfolio (K, H) to be *self-financing*, i.e. to satisfy the self-financing condition

$$K_tB_{t+1} + H_t \cdot S_{t+1} = K_{t+1}B_{t+1} + H_{t+1} \cdot S_{t+1}, \quad t \in \mathbb{T}, t < T. \tag{62}$$

In particular, while to invest we need to choose the values of the variables $K_t, H_t^1, \ldots, H_t^m$ for each $t \in \mathbb{T}, t < T$, not all these random variables are free ones, since they need to satisfy the T constraints of eq. (62). Since $B > 0$, we could consider eq. (62) as determining the value of K_{t+1} given the values of K_t, H_t, H_{t+1}, for every t, in which case the variables K_0 and $H_t, t \in \mathbb{T}, t < T$ would be chosen freely and the value of K_t for $t \in \mathbb{T} \cap (0, T)$ would be determined by induction: knowing the value of K_t we could then calculate that of K_{t+1} as

$$K_{t+1} = \frac{K_tB_{t+1} + (H_t - H_{t+1}) \cdot S_{t+1}}{B_{t+1}}. \tag{63}$$

Once calculated K by induction from the value of K_0 and H, we could compute the wealth as $K_tB_t + H_t \cdot S_t$. Doing this would be laborious; it would be much preferable to obtain instead formula for the wealth $V_t^{K,H}$ (and possibly also for K_t) as a function of K_0, H. This can be done, as follows.

We can express the change in value of the portfolio as the sum of the change in value of the investments in bonds, plus the change in value of the investments in stocks, as follows

$$V_{t_{i+1}} - V_{t_i} = K_{t_i}(B_{t_{i+1}} - B_{t_i}) + H_{t_i}(S_{t_{i+1}} - S_{t_i});$$

since $B_{t_{i+1}} - B_{t_i} = B_{t_i} r$, and $K_{t_i} B_{t_i} = V_{t_i} - H_{t_i} S_{t_i}$ is the value of the investments in bonds at time t_i, we have that

$$K_{t_i}(B_{t_{i+1}} - B_{t_i}) = K_{t_i} B_{t_i} r = (V_{t_i} - H_{t_i} S_{t_i}) r,$$

and so

$$V_{t_{i+1}} - V_{t_i} = H_{t_i}(S_{t_{i+1}} - S_{t_i}) + (V_{t_i} - H_{t_i} S_{t_i}) r \tag{64}$$

Summing over i, where $0 = t_0 < t_1 < \ldots < t_k$, we get

$$V_{t_k} - V_0 = \sum_{i=0}^{k-1} H_{t_i}(S_{t_{i+1}} - S_{t_i}) + (V_{t_i} - H_{t_i} S_{t_i}) r, \tag{65}$$

since the telescopic sum $\sum_{i=0}^{k-1} V_{t_{i+1}} - V_{t_i}$ equals $V_{t_k} - V_0$. The above equation, together with the identity $V_0 = K_0 B_0 + H_0 \cdot S_0$, allows us to express $V_t^{K,H}$ as a function only of K_0, H; moreover, since it shows that $V_{t_k} - V_0$ only depends on H, it suggests considering as variables the initial capital $x := V_0$ (and the number of assets $H_t, t \in \mathbb{T}, t < T$) instead of K_0, as this makes the formula for $V_t^{x,H}$ a little simpler and, most importantly, explicitly keeps track of x, which is a quantity of greater interest than K_0. If we work in discounted terms, we can express the wealth even more conveniently: indeed eq. (7) shows that the discounted equivalent of eq. (64) is the simpler equation

$$\overline{V}^{x,H}_{t_{i+1}} - \overline{V}^{x,H}_{t_i} = H_{t_i} \cdot (\overline{S}_{t_{i+1}} - \overline{S}_{t_i}),$$

which summed over i gives the neat formula

$$\overline{V}^{x,H}_t = x + (H \cdot \overline{S})_t, \quad t \in \mathbb{T} \tag{66}$$

where the 'discrete-time stochastic integral' (a.k.a. *martingale transform*) of H with respect to Y is defined as

$$(H \cdot Y)_t := \sum_{s=0}^{t-1} H_s \cdot (Y_{s+1} - Y_s), \quad t \in \mathbb{T} \tag{67}$$

where the $\cdot$ on the RHS (appearing after the term H_s) denotes the dot product in $\mathbb{R}^m$ (both H and Y are $\mathbb{R}^m$-valued processes).

Either way, using (K_0, H) or (x, H) as variables, and working in discounted terms or not, we can easily and directly express the wealth at time t, without having to first compute the number of bonds we need to buy. If we actually want to compute K_t, we can do so using the wealth V_t (or $\overline{V}_t$) at time t, which we just calculated, using the formula $K_t B_t = V_t - H_t \cdot S_t$, so that

$$K_t = \frac{1}{B_t}(V_t - H_t \cdot S_t) = \overline{V}_t - H_t \cdot \overline{S}_t.$$

So, from now on, we will always work in discounted terms and use (x, H) to describe a portfolio. Notice that, as H, B, S are adapted, so are $\overline{V}^{x,H}$ and K...as it should be the case: they should be known at time t!

3.3 Lecture 3, Arbitrage and arbitrage-free prices

In a multi-period model $(B_t, S_t)_{t\in\mathbb{T}}$ (where $S_t = (S_t^1, \dots, S_t^m)$), the notion of arbitrage is essentially unchanged: a trading strategy H (i.e. an adapted process with values in $\mathbb{R}^m$) is an *arbitrage* if $\overline{V}_T^{0,H} \geq 0$ a.s. and $\overline{V}_T^{0,H}$ is not a.s. $= 0$.

When we consider a derivative which has a payoff X_T only at time T, we need to determine the fair value X_t of X at all $t \in \mathbb{T}, t < T$. If $T = 1$ (one-period model) this means we only have to determine X_0, which is deterministic; if $T > 1$ (multi-period model) we have to determine $X_0, X_1, \dots, X_{T-1}$, where X_t will be known at time t; since normally we prefer to talk about processes as being defined for all $t \in \mathbb{T}$, we will use the following definition.

Definition 82. An adapted process $(P_t)_{t\in\mathbb{T}}$ is an *Arbitrage Free Price* for the derivative with payoff X_T at maturity T in the arbitrage-free market model $(B_t, S_t)_{t\in\mathbb{T}}$ if $P_T = X_T$ and the enlarged market $(B_t, S_t, P_t)_{t\in\mathbb{T}}$ is arbitrage-free.

From now on we will use the notation X instead of P to denote an AFP of X (we didn't do so in the above definition to be able to write $P_T = X_T$). Clearly the above definition when applied to a one-period model states that (X_0, X_1) is an AFP of X_1 (in the sense of definition 82) iff X_0 is an AFP of X_1 (in the sense of definition 40), so the two definitions essentially coincide. The following is the multi-period analogue of theorem 36.

Theorem 83. *In the linear multi-period market model $(B_t, S_t)_{t\in\mathbb{T}}$ of eq. (66), the Domination Principle holds $\iff$ there exists no-arbitrage.*

Proof. (Implication $\Longrightarrow$:) By definition, an arbitrage is an investment which, when compared[90] to the zero investment[91] violates the strict domination principle (for $t = 0$). (Implication $\Longleftarrow$:) If the law of one price fails, there are two portfolios $L = (x, G), M = (y, H)$, with values V^L, V^M given as in eq. (66), s.t. $V_T^L = V_T^M$ a.s. and yet V_t^L is not a.s. equal to V_t^M. We can[92] then assume w.l.o.g that $p := \mathbb{P}(\{V_t^L > V_t^M\}) > 0$. Consider then the portfolio $N = (0, I)$, where

$$I_s(\omega) := \begin{cases} 0 & \text{if } s < t \\ 0 & \text{if } s \geq t, \quad \omega \in \{V_t^L < V_t^M\} \\ (H_s - G_s)(\omega) & \text{if } s \geq t, \quad \omega \in \{V_t^L \geq V_t^M\}. \end{cases} \tag{68}$$

Clearly[93] I is adapted. To show that I is an arbitrage, use that, if $s \geq t, \omega \in \{V_t^L \geq V_t^M\}$,

$$\overline{V}_{s+1}^N - \overline{V}_s^N = I_s \cdot (\overline{S}_{s+1} - \overline{S}_s) = (H_s - G_s) \cdot (\overline{S}_{s+1} - \overline{S}_s) = (\overline{V}_{s+1}^M - \overline{V}_s^M) - (\overline{V}_{s+1}^L - \overline{V}_s^L),$$

[90] Considering the 0 investment as the first investment, an the arbitrage as the second one.
[91] i.e. to having no capital and doing nothing.
[92] Otherwise $\mathbb{P}(\{V_t^M > V_t^L\}) > 0$, so we can just reverses the roles of L and M.
[93] Since one can write the value of I_s as a function of quantities that are known at time s.

and otherwise $\overline{V}^N_{s+1} - \overline{V}^N_s = 0$, and so summing over s we get that

$$\overline{V}^N_T - \overline{V}^N_0 = 1_{\{V^L_t \geq V^M_t\}}((\overline{V}^M_T - \overline{V}^M_t) - (\overline{V}^L_T - \overline{V}^L_t)),$$

and since $\overline{V}^N_0 = 0$ we get

$$\overline{V}^N_T = 1_{\{V^L_t \geq V^M_t\}}(\overline{V}^M_T - \overline{V}^L_T) + 1_{\{V^L_t \geq V^M_t\}}(\overline{V}^L_t - \overline{V}^M_t)). \tag{69}$$

Since we assumed $V^L_T = V^M_T$ a.s., we get $\overline{V}^N_T = 1_{\{V^L_t \geq V^M_t\}}(V^L_t - V^M_t)$, which is ≥ 0 a.s. and > 0 with probability $p > 0$, so I is an arbitrage.

Analogously, if the strict domination principle fails there are $L = (x, G), M = (y, H)$ s.t. $V^L_T \leq V^M_T$ a.s., the equality $V^L_T = V^M_T$ does not hold a.s., and yet either $V^L_t \leq V^M_t$ does not hold a.s. (i.e. $\mathbb{P}(\{V^L_t > V^M_t\}) > 0$), or $V^L_t = V^M_t$ a.s.. It is easy to see that also in this case I defined in eq. (68)) is an arbitrage, as eq. (69) expresses $\overline{V}^N_T$ as the sum of two random variables which are both ≥ 0 a.s., and the second (*resp. first*) one of which is not a.s. $= 0$ if $\mathbb{P}(\{V^L_t > V^M_t\}) > 0$ (*resp. if* $V^L_t = V^M_t$ *a.s.*). □

Theorem 83 establishes the link between the absence of arbitrage (a property about what happens over the time span from 0 to T) and the domination principle (a property about what happens over the time span from t to T, for any t). It is then not surprising that one can reduce the study of existence of arbitrage in multi-period models to one-period models; this is indeed the message of the following theorem.

Theorem 84. *There exists an arbitrage in the multi-period model* $(B_t, S_t)_{t \in \mathbb{T}}$ *if and only if there exists a* $s \in \mathbb{T}, s < T$ *such that there exists an arbitrage in the one-period model* $(B_t, S_t)_{t=s,s+1}$.

Proof. We only prove the simpler implication $\Longleftarrow$. Let the $\mathcal{F}_s$-measurable random variable A_s be an arbitrage in the one-period model $(B_t, S_t)_{t=s,s+1}$, so that $W := A_s \cdot (\overline{S}_{s+1} - \overline{S}_s) \geq 0$ a.s. and it not a.s. $= 0$. Consider then the strategy $(H_t)_{t \in \mathbb{T}, t<T}$ given by: $H_t = 0$ if $t \neq s$, $H_s = A_s$, which corresponds to doing nothing before time s, at s 'buying' A_s shares ('borrowing' from the bank), at time $s + 1$ 'selling' those A_s shares ('depositing[94]' money in the bank) and then doing nothing, so that until maturity we hold 0 shares. Then H is clearly an arbitrage in $(B_t, S_t)_{t \in \mathbb{T}}$, since it is adapted and its payoff at time T is $V^{0,H}_T = W \Pi^{T-1}_{t=s+1}(1 + R_t)$, which is ≥ 0 a.s. and it not a.s. $= 0$, since such is W, and $R > -1$. □

3.4 Lecture 4, The multi-period binomial model

Let us now consider the multi-period binomial model. Just like in the one-period setting, we consider only one underlying, whose price S at each time can jump to only two possible values. Correspondingly, we can think of a coin being tossed at every time, and determining the value to which S will jump to. To describe this, we can consider

[94] As usual if $A_s < 0$ we are actually selling $-A_s$ shares and depositing money in the bank, etc.

a process X representing the results of the coin toss: if at time n is we get Heads the process takes the value $X_n = H$, and if Tails then $X_n = T$. From now on we will denote maturity with N, not with T, and the trading strategy with G, not with H to avoid confusion (as the symbols T, H are used for Tails and Heads). We can then build the N-period binomial model on the following filtered probability space $(\Omega, \mathcal{A}, \mathcal{F}, \mathbb{P}_{\mathbf{p}})$:

1. $\Omega = \Omega_N := \{H, T\}^N$, i.e. $\omega = (\omega_1, \cdots, \omega_N) \in \Omega$ with $\omega_n \in \{H, T\}$ $\forall n$.

2. $\mathcal{A} := \mathcal{P}(\Omega) :=$ all subsets of Ω.

3. $\mathcal{F} = (\mathcal{F}_n)_{n=0}^N$ is the natural[95] filtration of X, where $X_n(\omega) := \omega_n$ for any $\omega = (\omega_1, \cdots, \omega_N) \in \Omega$, i.e. X_n is the projection from the product space $\Omega = \{H, T\}^N$ to the n^{th}-component $\{H, T\}$ in that product.

4. Given $p \in (0, 1)$, define the probability $\mathbb{P}_p$ on $\{H, T\}$ by setting $\mathbb{P}_p(\{H\}) := p$, $\mathbb{P}_p(\{T\}) := 1 - p$. Given $\mathbf{p} = (p_n)_{n=1}^N \in (0, 1)^N$, define $\mathbb{P}_{\mathbf{p}}$ on $\Omega = \{H, T\}^N$ as $\Pi_{n=1}^N \mathbb{P}_{p_n}$, i.e.

$$\mathbb{P}_{\mathbf{p}}(\{(\omega_1, \cdots, \omega_N)\}) = \Pi_{n=1}^N \mathbb{P}_{p_n}(\{\omega_n\}) \text{ for any } \omega = (\omega_1, \cdots, \omega_N) \in \Omega,$$

   and then $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$ for all $A \in \mathcal{A}$. For example, if $N = 2$ then[96]

$$\mathbb{P}(HH) = p_1 p_2, \quad \mathbb{P}(HT) = p_1(1-p_2), \quad \mathbb{P}(TH) = (1-p_1)p_2, \quad \mathbb{P}(TT) = (1-p_1)(1-p_2).$$

   Notice that $\mathbb{P}_{\mathbf{p}}$ is a probability, since clearly $\mathbb{P}_{\mathbf{p}}(\{\omega\}) \geq 0$ for every $\omega \in \Omega$ (because the product of positive numbers is positive), and $\mathbb{P}_{\mathbf{p}}(\Omega) = 1$ (because[97] $1^N = 1$).

In what follows we will always consider processes on $(\Omega, \mathcal{A}, \mathcal{F}, \mathbb{P}_{\mathbf{p}})$. Notice that a process Y is adapted if, for each n, Y_n is a function of $X(n) := (X_k)_{k \leq n}$, or equivalently if $Y_n(\omega)$ it depends *only* on $\omega(n) := (\omega_1, \ldots, \omega_n)$: it does not depend on $(\omega_{n+1}, \ldots, \omega_N)$. In this case we may write (slightly improperly) $Y_n(\omega(n))$ to mean the constant value $Y_n(\omega')$ which Y_n takes on the set $\{\omega' : X(n)(\omega') = \omega(n)\} =: \{X(n) = \omega(n)\}$. Analogously, we may improperly[98] write $\{HHT\}$ to mean

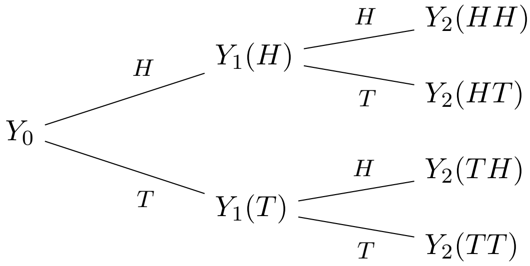
$$\{X(3) = HHT\} := \{\omega' : X(3)(\omega') = HHT\} = \{\omega' \in \{H, T\}^N : \omega'_1 = H, \omega'_2 = H, \omega'_3 = T\}.$$

We identify adapted processes with binary trees, as we now illustrate for $N = 2$. If Y is an adapted process, i.e. Y_0 is constant and $Y_1(\omega)$ only depends on the value of ω_1, we can write its values on a binary tree as follows

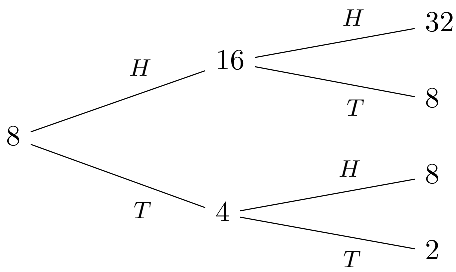
[95] i.e. $\mathcal{F}_n := \sigma((X_k)_{k \leq n})$ for all n.
[96] The expression $\mathbb{P}(HH)$ (and similar ones) is a convenient abbreviation for $\mathbb{P}(\{HH\})$.
[97] For example if $N = 2$ we get

$$\mathbb{P}_{\mathbf{p}}(\Omega_N) = p_1 p_2 + p_1(1 - p_2) + (1 - p_1)p_2 + (1 - p_1)(1 - p_2) = [p_1 + (1 - p_1)][p_2 + (1 - p_2)] = 1^2 = 1.$$

[98] This is improper since $\{HHT\}$ is a subset of $\{H, T\}^3$, not of $\{H, T\}^N$.



Conversely, writing on a binary tree some values, e.g.



identifies a unique adapted process Y which takes those values on those nodes, given by

$$Y_0 = 8, Y_1(H) = 16, Y_1(T) = 4, Y_2(HH) = 32, Y_2(HT) = 8 = Y_2(TH), Y_2(TT) = 2.$$

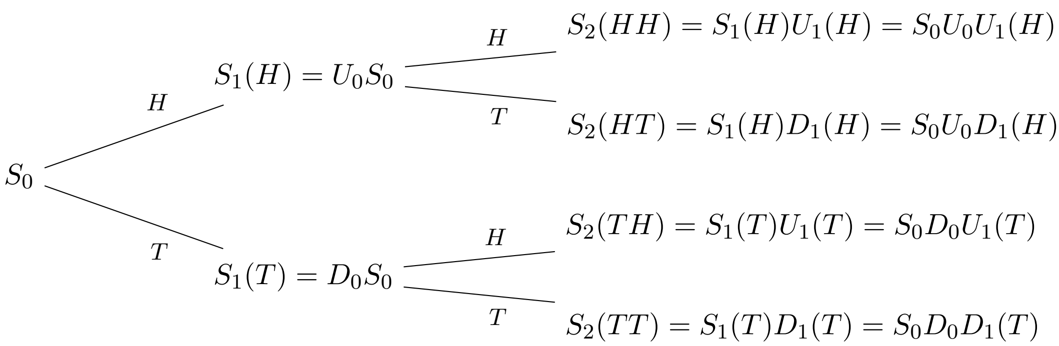
On $(\Omega, \mathcal{A}, \mathcal{F}, \mathbb{P}_{\mathbf{p}})$ we build the binomial market model (B, S) by asking that:

- the bond price process B is given by $B_n = B_0 \Pi_{k=0}^{n-1}(1 + R_k)$ for each $n \in \mathbb{T}$, where the interest rate process $R > -1$ is $\mathcal{F}$-adapted.
- the price of the underlying is given by

$$S_{n+1}(\omega) = \begin{cases} (S_n U_n)(\omega) & \text{if } X_{n+1}(\omega) = H, \\ (S_n D_n)(\omega) & \text{if } X_{n+1}(\omega) = T, \end{cases} \tag{70}$$

  where the up and down factors U_n, D_n are $\mathcal{F}$-adapted processes s.t. $0 < D < U$.

For example, in the 2-period setting, the stock price is given by



Often we take the processes R, U, D to be constants r, d, u (i.e. deterministic and not depending on time): this is the classic binomial model as considered by most authors.

The following theorem follows immediately from theorem 84, and the fact that in the one-period binomial model with time index $\{n, n+1\}$ there is no arbitrage iff $D_n < 1 + R_n < U_n$.

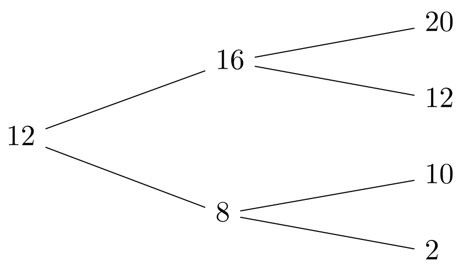
Theorem 85. *The multi-period binomial model is arbitrage-free* $\iff D < 1 + R < U$.

3.5 Lecture 5, Example of pricing in the multi-period binomial model

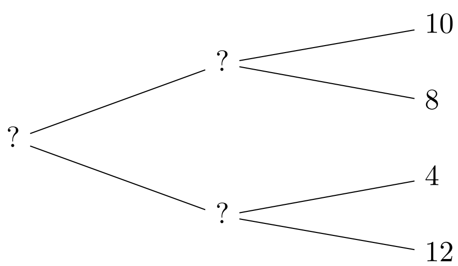
Let us now show with an example how in the multi-period binomial model we can price any derivative by replication, using the following method, called *backward induction*; this shows in particular that the model is *complete*.

Using our knowledge of the one-period binomial model, we can determine the replicating strategy G_{n-1} and the wealth $V_{n-1}^{x,G}$ at time $n-1$ from the (known) wealth $V_n^{x,G}$ at time n. Thus, if we start from $n = N$, at which time the wealth $V_N^{x,G}$ is known and must equal the final payoff X_N to be replicated, we can keep decreasing n until $n = 1$, at which point we have computed H_n and $V_n^{x,G}$ for every n, and in particular we also computed $x = V_0^{x,G}$.

Example 86. Consider the 2-period binomial model with stock price S described by this tree

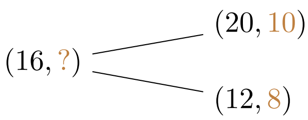


Consider now a derivative with maturity 2 whose payoff $Y_2 = f(S_0, S_1, S_2)$ in the present model is given by the following random variable:



Let us work in the model where the interest rate R equals 0 (i.e. $R_0 = R_1 = 0$ a.s.), which is only a very mild simplification[99].

To price Y_2, imagine that we are at time 1, that the first coin has been tossed and resulted in Heads, i.e. $\omega_1 = H$. Then it is like if we were in a binomial model with 1 time period, i.e. the values of stock and derivative (S, Y) are



Thus, to compute the price $Y_1(H)$ we can apply our knowledge of the one-period binomial model: we can replicate any payoff in a unique way, and the time 1 value of the portfolio replicating Y_2 knowing that $\omega_1 = H$ is found using the risk neutral pricing formula. To do so, we compute the up and down factors

$$U_1(H) = \frac{S_2(HH)}{S_1(H)} = \frac{20}{16} = \frac{5}{4}, \qquad D_1(H) = \frac{S_2(HT)}{S_1(H)} = \frac{12}{16} = \frac{3}{4}$$

and from them we compute the 'risk-neutral probability given $\{X_1 = H\}$'

$$\tilde{P}_1(H) := \frac{1 + R_1 - D_1}{U_1 - D_1}(H) = \frac{1 + 0 - 3/4}{5/4 - 3/4} = \frac{1}{2}, \qquad \tilde{Q}_1(H) := 1 - \tilde{P}_1(H) = \frac{1}{2}$$

and we use them to evaluate the expectation under the 'risk-neutral measure given $\omega_1 = H$' by taking $\omega_1 = H$ in the formula

$$V_1^{x,G}(\omega_1) = \mathbb{E}^{\mathbb{Q}}\left[\frac{Y_2}{1+R_1}\middle| X_1 = \omega_1\right] := \frac{\tilde{P}_1(\omega_1)Y_2(\omega_1 H) + \tilde{Q}_1(\omega_1)Y_2(\omega_1 T)}{1 + R_1(\omega_1)} \tag{71}$$

which gives

$$V_1^{x,G}(H) = \frac{10+8}{2} = 9.$$

Since $V_1^{x,G}(H) = 9$ equals the value (at time 1 and if $\omega_1 = H$) of a portfolio replicating replicating Y_2, by the law of one price $Y_1(H) = 9$. we can also compute the value $G_1(H)$ of the replicating strategy G at time 1 when $\omega_1 = H$ by taking $\omega_1 = H$ in the delta-hedging formula

$$G_1(\omega_1) := \frac{X_2(\omega_1 H) - X_2(\omega_1 T)}{S_2(\omega_1 H) - S_2(\omega_1 T)}, \tag{72}$$

which gives

$$G_1(H) = \frac{10-8}{20-12} = \frac{2}{8} = \frac{1}{4}.$$

[99]Otherwise we would have to work in discounted terms, i.e. draw the tree of the discounted stock price $\overline{S}$ and payoff $\overline{Y}_2$, and price $\overline{Y}_2$ in the binomial model with price $\overline{S}$ and interest rate 0.

Analogously we can compute what happens when $\omega_1 = T$ and find

$$U_1(T) = \frac{5}{4}, \quad D_1(T) = \frac{1}{4}, \quad \tilde{P}_1(T) = \frac{3}{4}, \quad \tilde{Q}_1(T) = \frac{1}{4}$$

and then the risk neutral pricing formula and the delta-hedging formula give

$$V_1^{x,G}(T) = \frac{3}{4} \cdot 4 + \frac{1}{4} \cdot 12 = 6, \quad G_1(T) = \frac{4-12}{10-2} = -1.$$

Remark 87. The quantity $\mathcal{R}_1^2(Y_2) = \mathbb{E}^{\mathbb{Q}}[\frac{Y_2}{1+R_1}|X_1 = \omega_1]$ which appears in eq. (71) represents the value at time 1 which, appropriately invested, replicates the payoff Y_2 at time 2; in other words, $\mathcal{R}_1^2(Y_2)$ represents the value at time 1 of having the amount Y_1 at time 2. Analogously one can consider the operator $\mathcal{R}_k^n$ which, given a payoff Y_n at time n, returns its value $Y_k = \mathcal{R}_k^n(Y_n)$ at time $k \leq n$. Clearly $\mathcal{R}_i^j \mathcal{R}_j^k = \mathcal{R}_i^k$ for $i \leq j \leq k$, and so $\mathcal{R}_i^k = \mathcal{R}_i \mathcal{R}_{i+1} \ldots \mathcal{R}_{k-1}$ for $\mathcal{R}_n := \mathcal{R}_n^{n+1}$; thus, to compute $\mathcal{R}_i^k$ it is enough to be able to compute the *rollback operator*

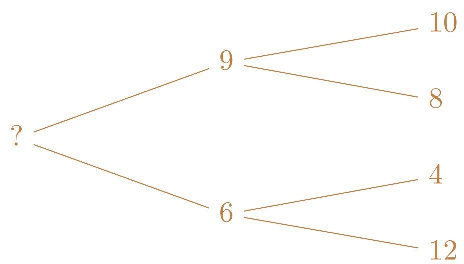
$$\mathcal{R}_n(Y_{n+1})(\omega_1, \ldots, \omega_n) = \mathbb{E}^{\mathbb{Q}}\left[\frac{Y_{n+1}}{1+R_n}\middle| X_1 = \omega_1, \ldots, X_1 = \omega_n\right], \tag{73}$$

which we write more succinctly as

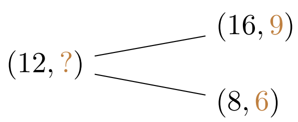
$$\mathcal{R}_n(V_{n+1}) = \mathbb{E}^{\mathbb{Q}}\left[\frac{V_{n+1}}{1+r}\middle| X_1, \ldots, X_n\right],$$

and whose definition is entirely analogous to that of $\mathcal{R}_2$ which appears in eq. (71); we will later see how more generally we can formally define $\mathbb{E}^{\mathbb{Q}}[Y|X]$, where Y is a random variable, X a random vector, and $\mathbb{Q}$ a probability.

We can now add the values of $Y_1(H) = V_1^{x,G}(H) = 9, Y_1(T) = V_1^{x,G}(T) = 6$ in the tree of Y as follows



To conclude, we just need use the one-period binomial model on the time $\{0, 1\}$ to price and hedge a derivative with payoff Y_1. In other words, we consider the values of stock and derivative (S, Y) to be



from which we compute

$$U_0 = \frac{S_1(H)}{S_0} = \frac{16}{12} = \frac{4}{3}, \quad D_0 = \frac{S_1(T)}{S_0} = \frac{8}{12} = \frac{2}{3}$$

and the 'risk-neutral probability'

$$\tilde{P}_0 = \frac{1 + r_0 - d_0}{u_0 - d_0} = \frac{1 - 2/3}{4/3 - 2/3} = \frac{3-2}{4-2} = \frac{1}{2}, \quad \tilde{Q}_0 = 1 - \tilde{P}_0 = \frac{1}{2}$$

and finally the initial value of a portfolio replicating Y_1 (and thus Y_2) is

$$x = V_0^{x,G} = \frac{\tilde{P}_0 X_1(H) + \tilde{Q}_0 X_1(T)}{1 + r_0} = \frac{1}{2}(9+6) = \frac{15}{2}$$

and the initial value of the replicating strategy is

$$G_0 = \frac{V_1^{x,G}(H) - V_1^{x,G}(T)}{S_1(H) - S_1(T)} = \frac{9-6}{16-8} = \frac{3}{8}.$$

Remark 88. As the reasoning used to work out the above example 86 makes clear, the N-period binomial model is complete, i.e. for any payoff X_N at maturity N one can find a replicating portfolio, i.e. a (x, G) s.t. $V_N^{x,G} = X_N$.

Given $\omega_1 \in \{H, T\}$, we have interpreted the quantities $\tilde{P}_1(\omega_1)$ and $\tilde{Q}_1(\omega_1)$ as 'risk-neutral probabilities, of $\{X_2 = H\}$ and of $\{X_2 = T\}$, given $\{X_1 = \omega_1\}$'; analogously we interpreted $\tilde{P}_0$ and $\tilde{Q}_0$ as 'risk-neutral probabilities of $\{X_1 = H\}$ and of $\{X_1 = T\}$'. This suggests that we should indeed define what is means to be a risk-neutral probability $\mathbb{Q}$ on Ω_N, and show that there exists a unique such $\mathbb{Q}$ which, for all $\omega_1 \in \{H, T\}$, satisfies

$$\mathbb{Q}(H) = \tilde{P}_0, \quad \mathbb{Q}(T) = \tilde{Q}_0, \quad \mathbb{Q}(X_2 = H | X_1 = \omega_1) = \tilde{P}_1(\omega_1), \quad \mathbb{Q}(X_2 = T | X_1 = \omega_1) = \tilde{Q}_1(\omega_1).$$

Week 7

3.6 Lecture 1, Derivatives paying a cashflow

So far we considered only derivatives which provide a payoff P_T at only one, deterministic (i.e. non-random) time $T = N \in \mathbb{N}$. We will now discuss derivatives which provide a payoff P_t at any time $t = 0, 1, \dots, T$, where $T \in \mathbb{N}$ is still called *expiry* (or *maturity*). The same approach applies when the derivative has a payoff at time $t \in \mathbb{N}$, as long as the infinite sums which come up in the computations converge. The process $P = (P_t)_{t=0,\dots,T}$ is called the *cashflow* of the derivative, and must be *adapted* (to the underlying filtration $\mathcal{F} = (\mathcal{F}_t)_t$), since the amount that a derivative pays at time t should be known at time t. As usual, though we called it 'payoff', the quantity $P_t \in \mathbb{R}$ does not need to be positive. When $P_t = 0$ at all times $t < T$, then we are back to the previous case of a payoff only at time T, so the derivatives paying a cashflow generalise those that have a payoff at one time only.

To price derivative with a cashflow P one can work by replication, as usual. Indeed, a derivative with a cashflow P can be seen as the sum over $n = 0, \dots, N$ of derivatives with payoff P_n at expiry n, so to replicate and price the cashflow one just needs to replicate and price each derivative with only one payoff, and then sum up the results. Of course the underlying reason why this works is our crucial assumption of linearity of the map that given a portfolio outputs its value. Thus, denote with P_k^n the value at time $k \leq n$ of the derivative which only has payoff P_n at time n, and with H_k^n the number of shares one should hold at time k to replicate it. Notice that the value of P_k^n is defined only for $k \leq n$, since the derivative has expiry n, and analogously H_k^n is defined only for $k \leq n - 1$.

Clearly at any time we can only ask to replicate the *future* cash flows of a derivative, so the replicating strategy H_k (of the derivative which pays the cashflow P) at time k must just replicate the *future* cash flows. Analogously, the value V_k at time k of the derivative which pays the cashflow P equals the present value of its *future* cash flows. Thus, in formulas

$$H_k = \sum_{n=k+1}^{N} H_k^n, \qquad V_k = \sum_{n=k}^{N} D_k^n. \tag{74}$$

We already know how to compute such P_k^n and H_k^n, as this deals only with a derivative with payoff only at time n; thus, eq. (74) allows us to price the whole cashflow P.

3.7 Lecture 2, Derivatives with random maturity

We now want to consider derivatives with random maturity. We model such maturity using the concept of random time τ, and of evaluating a process P at a random time, which we now define. Throughout we consider a set $I \subseteq \mathbb{R}$, which represents the *time-index*; when working in discrete-time we take $I = \mathbb{N}$ (or sometimes $I = \{0, 1, \dots, N\}$); when working in continuous time one analogously considers $I = \mathbb{R}_+$ (or $I = [0, T]$).

Definition 89. A random variable with values in $I \cup \{\infty\}$ is called a *random time.*

We allow the random time to take the value ∞, to allow for the possibility that the derivative pays nothing at all (this happens on the event $\{\tau = \infty\}$).

Given a process $P = (P_i)_{i \in I}$, and a random time τ (all defined on the same probability space), we define P_τ as the random variable which takes the value P_t on the event $\{\tau = t\}$, for all $t \in I$, and takes the value 0 on the event $\tau = \infty$.

A derivative with random maturity τ is then simply a derivative with cashflow P, which has a non-zero payoff P_τ only at one *random time* τ (i.e. $P_t(\omega) = 0$ whenever $t \neq \tau(\omega)$).

For a derivative paying a cashflow P which is non-zero at at most one time τ, to know whether $\tau(\omega) \leq n$ (i.e. that the derivative has already provided its payoff before time n), it is enough to know what happens up to time n; we formalise this intuitive concept using the following definition.

Definition 90. A random time τ is called a *stopping time* if $\{\tau \leq t\}$ is $\mathcal{F}_t$-measurable for all $t \in I$.

Notice that if $I = \mathbb{N}$ then τ is a stopping time iff $\{\tau = t\}$ is $\mathcal{F}_t$-measurable for all $t \in I$; not so for $I = \mathbb{R}$, since in this case the event $\{\tau \leq t\}$ is given by the *uncountable* union of the events $\{\tau = s\}, s \leq t$ (which may be not measurable even if each of these events is measurable).

For an example of a stopping time with $I = \mathbb{N}$, consider an arbitrary derivative paying a cashflow, modelled as the adapted process P. Then the 1^{st} time of payment

$$\tau_1 := \tau_1(P) := \inf\{t \in I : P_t \neq 0\}$$

of P is a stopping time, as it follows from the identity

$$\{\tau_1 = t\} = \{P_t \neq 0\} \cap (\cap_{s=0}^{t-1}\{P_s = 0\}),$$

which expresses $\{\tau_1 = t\}$ as a combination of events which are $\mathcal{F}_t$-measurable (since P is adapted to $\mathcal{F}$). As usual we are using here the definition $\inf \emptyset := \infty$; in particular τ_1 has values in $\mathbb{N} \cup \{\infty\}$, and $\tau_1(\omega) = \infty$ holds if and only if $P_t(\omega) = 0$ for all $t \in \mathbb{N}$. Analogously the k^{th} time of payment $\tau_k := \tau_k(P) := \inf\{t > \tau_{k-1} : P_t \neq 0\}$ of P is a stopping time for any $k = 1, 2, \ldots$, as it follows by induction using the identity

$$\{\tau_{k+1} = t\} = \cup_{s=k-1}^{t-1}\Big(\{\tau_k = s\} \cap \{P_t \neq 0\} \cap (\cap_{u=s+1}^{t-1}\{P_u = 0\})\Big).$$

Clearly saying that $P_t = 0$ for all $t \neq \tau$ for some stopping time τ is equivalent to saying that 2^{nd} time of payment $\tau_2(P)$ of P equals the constant ∞. In this case the derivative can be considered as having just one payoff P_τ at one time τ, but with τ being random (more precisely, τ will be the stopping time $\tau_1(P)$).

This suggests that it is possible to price it by replication up to time τ (which should be considered as a random expiry), in the same way as in the case in which the expiry was a constant. Indeed, though one could price and replicate derivatives with random expiry with the same method as for those paying any cashflow, it is simpler and quicker to instead set up and solve the replication equation only up to time τ. In this case the replicating strategy H and the value V of the derivative are defined only up to time $\tau - 1$ and τ respectively (as is the case when τ is a constant); alternatively, such values (i.e. the values of G_t for $t > \tau - 1$, and of X_t for $t > \tau$) could be considered as being 0.

Any derivative paying a cashflow can be seen as sum of derivatives paying only at one time. If we write such decomposition in the case of a derivative with random expiry we get the decomposition $P_\tau = \sum_{n=0}^{N} P_n 1_{\{\tau=n\}}$, where $P_n 1_{\{\tau=n\}}$ only pays at time $n \leq N$. The fact that $P_t = 0$ for all but at most one values of t corresponds to the fact that the payoff $P_n 1_{\{\tau=n\}}$ is not 0 for at most one n.

In most examples, τ is the *hitting time* of a set C relative to some process X, which is either the price S of the underlying, or some functional thereof (e.g. the running maximum $X_t := \sup_{u \leq t} S_u$). In other words, normally τ is the first time X hits C, i.e. $\tau := \inf\{k \geq 0 : X_k \in C\}$. Clearly such τ is a stopping time if X is adapted, as it follows from the identity

$$\{\tau = t\} = \{X_t \in C\} \cap (\cap_{s=0}^{t-1} \{X_s \notin C\}).$$

Let us now consider an explicit example.

Example 91. In the binomial model $N = 2, r = 0, S_0 = 4, u = 2, d = \frac{1}{2}$, find the price X the option which pays $(S_N - 3)^+$ at time N if S never crosses the barrier $U = 6$ (i.e. $S_t \leq 6$ for all t), and instead pays a different amount, called *rebate*, which in the present example we assume to equal 2, at the time

$$\sigma := \inf\{k = 0, 1, 2 : S_k \geq U\}$$

that S crosses the barrier U.

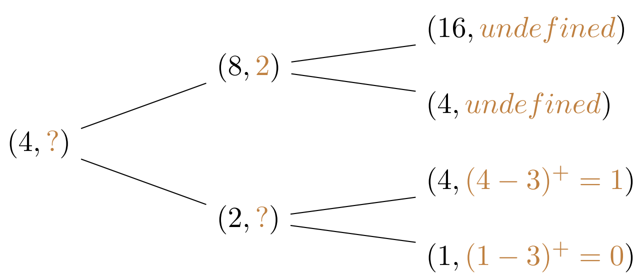


Figure 2: Tree of (S, X).

Let us now compute the price X of the derivative up to its expiry. Notice that the time of expiry is not σ, since by definition $\sigma = \infty$ if (and only if) S never crosses the barrier $u = 6$, and in this case the derivative has a payoff at time $N = 2$. Thus, the expiry of this derivative is time $\tau := \sigma \wedge N$.

ω	HH	HT	TH	TT
$\sigma(\omega)$	1	1	∞	∞
$\tau(\omega)$	1	1	2	2

Figure 3: Values of σ and τ.

To compute X we work as usual by backward induction in each branch of the binomial tree, starting from time τ. Thus, we first consider the branch emanating from $\omega_1 = T$

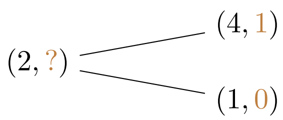


Figure 4: Branch of tree of (S, X) from $\omega_1 = T$.

Since $r = 0, \tilde{p} := \frac{(1+r)-d}{u-d} = \frac{1-\frac{1}{2}}{2-\frac{1}{2}} = \frac{1}{3}$, the RNPF gives

$$X_1(H) = \frac{1}{3} \cdot 1 + \frac{2}{3} \cdot 0 = \frac{1}{3}.$$

Given that for $\omega_1 = H$ we have $\tau(\omega) = 1$, we already know the value of $X_1(\omega) = X_\tau(\omega)$, which by definition equals the rebate value 2. Thus having computed $X_1(H) = \frac{1}{3}$ and knowing $X_1(T) = 2$, we can now compute X_0 by backward induction.

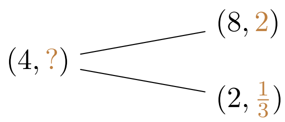


Figure 5: Root of tree of (S, X).

Applying the RNPF gives

$$X_0 = \frac{1}{3} \cdot 2 + \frac{2}{3} \cdot \frac{1}{3} = \frac{8}{9}.$$

Finally, the delta-hedging formula gives the replicating strategy

$$G_1(T) = \frac{X_2(TH) - X_2(TT)}{S_2(TH) - S_2(TT)} = \frac{1-0}{4-1} = \frac{1}{3},$$

$$G_0 = \frac{X_1(H) - X_1(T)}{S_1(H) - S_1(T)} = \frac{2-\frac{1}{3}}{8-2} = \frac{5}{18}.$$

Thus, we have determined the replicating strategy $G = (G_t)_{t \leq \tau - 1}$ and the value $X = (X_t)_{t \leq \tau}$ of the derivative.

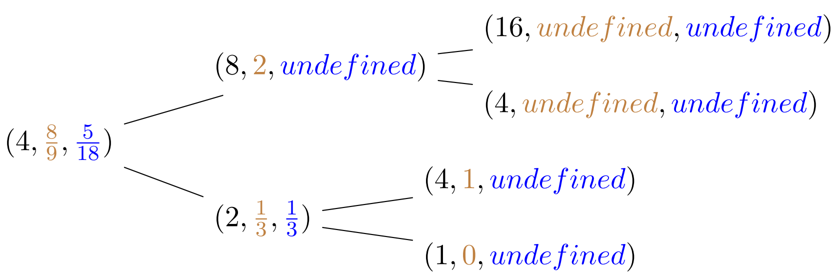


Figure 6: Tree of (S, X, G).

3.8 Lecture 3, Chooser options

Many derivatives offer the holder (buyer) the right to make choices during the lifetime of the contract. For example, if at time 0 I buy a *chooser option*, I get to choose, at time t, whether I will receive the payoff A at time $u \geq t$, or the payoff B at time $v \geq t$, where A and B can depend on t. More generally, the choice could be between many alternatives $(A^i)_{i \in I}$, instead of just two, and the corresponding times of payment $(\tau^i)_{i \in I}$, as well as the time τ at which we choice has to be made, could be stopping times (and $(A^i)_{i \in I}$ could depend on τ). Let us consider several important sub-cases of such derivatives.

A most important example is that of an *American* call option. If at time 0 I buy one then, at any time t, based on the information available at that time, I get to choose whether I want to exercise my option and receive the payment $(S_t - K)^+$ (after which I can no longer use the option), or instead wait, in the hope that exercising the derivative at a future time will lead to a preferable outcome. More formally, if I buy the American call option I have the right to choose a stopping time τ at which I will receive the payoff $(S_\tau - K)^+$. This is a special case of the above chooser option, where one considers as the only possible payment $(S_\tau - K)^+$, which however depends on the time τ of choice, and which is paid immediately when the choice is made. Notice that if it would be equivalently for me to consider buying the option which gives payoff $S_\tau - K$ at a time τ of my choose; indeed, in this case I would never choose to get paid at a time at which the payment is negative (if they payment would be negative at all times, i.e. $S_t - K < 0$ for all t, I would just choose not to exercise the option, i.e. choose $\tau = \infty$).

Analogously, if I buy a *Bermudan* call option, I get to choose the (stopping) time τ at which I will receive the payoff $(S_\tau - K)^+$, but only among those τ which take values in a set D of possible exercise dates.

We now explain how to compute the smallest price C at which one should be willing to sell a chooser option. This is calculated as being the smallest possible super-replication price, so that the one call sell the chooser and hedge it, no matter what choice the buyer makes. To compute such price, one can work by backward induction (applying the rollback operator, see remark 87) to compute the values $(A^i_t)^{i \in I}$ at any time $t \in [s, T]$ of each choice. At the time of choice s the option holder chooses the best alternative, so $C_s := \max((A^i_s)^{i \in I})$. Then one can compute C_r by backward induction for $r \leq s$.

A bit more complicated is the case in which the time $s = \sigma$ at which the choice is made is not fixed, but can itself be chosen (among some family $\mathcal{T}$ of stopping times), as is the case for the American option.

Example 92. In the binomial model $N = 2, r = 0, S_0 = 4, u = 2, d = \frac{1}{2}$, find the price C_0 at which you are willing to sell the option which gives the buyer the right to choose at time 1 whether to receive $A_2 := (S_2 - 5)^+$ at time 2, or $B_1 := (S_1 - 5)^+$ at time 1.

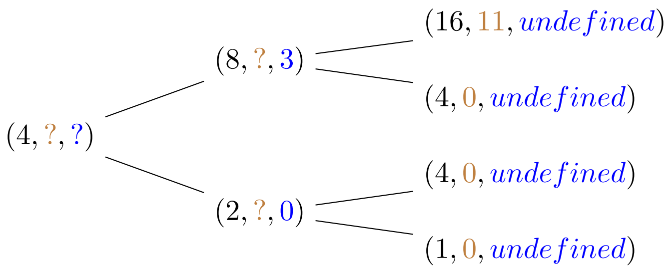


Figure 7: Tree of (S, A, B).

Let us now compute by backward induction the values of the possible choices at the time of choice $s = 1$.

Figure 8: Branches of tree of (S, A) from $\omega_1 = H$, and from $\omega_1 = T$.

Let us now apply the RNPF to the one-period model between $t = 1$ and $t = 2$, to obtain the replication price at $t = 1$. Notice that this price will depend on the information available at time 1 (i.e. the value of the first coin toss), since so does the model at time 1, so for now we will write the RNFP formally as $A_1 = \mathbb{E}_1^{\mathbb{Q}}\left[\frac{A_2}{1+r}\right]$ (later we will define what $\mathbb{E}_1^{\mathbb{Q}}$ actually means; for now, it is enough that it is clear to you how to compute it). Since $r = 0, \tilde{p} = \frac{1}{3}$, , the RNPF gives

$$A_1(H) = \frac{1}{3} \cdot 11 + \frac{2}{3} \cdot 0 = \frac{11}{3}, \quad A_1(T) = \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 0 = 0.$$

We can then compute the value C_1 of the chooser at time 1 as $C_1 = \max(A_1, B_1)$. Since $A_1(H) > B_1(H)$ the option holder should choose A_2 if $\omega_1 = H$. Since $A_1(T) = B_1(T)$, the option holder is indifferent between A_2 and B_1 if $\omega_1 = T$.

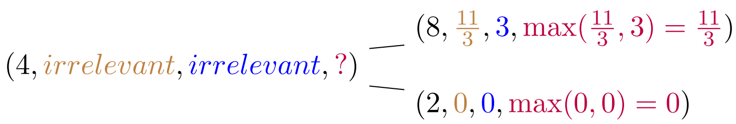


Figure 9: Root of tree of (S, A, B, C).

Finally, to compute C_0 we proceed again by backward induction from C_1 (so we do not need to compute the values A_0, B_0): since $r = 0, \tilde{p} = \frac{1}{3}$, the RNPF gives

$$C_0 = \frac{1}{3} \cdot \frac{11}{3} + \frac{2}{3} \cdot 0 = \frac{11}{9}.$$

Finally, the hedging strategy G is given by the delta-hedging formula

$$G_0 = \frac{C_1(H) - C_1(T)}{S_1(H) - S_1(T)} = \frac{\frac{11}{3} - 0}{8 - 2} = \frac{11}{18}.$$

If at time 1 the option holder chooses A_2, the option seller should hold

$$G_1(\omega_1) = \frac{A_2(\omega_1 H) - A_2(\omega_1 T)}{S_2(\omega_1 H) - S_2(\omega_1 T)}, \qquad \text{so } G_1(H) = \frac{11}{12}, G_1(T) = 0.$$

If instead at time 1 the holder chooses B_1, then the chooser expires, so no more hedging is needed, i.e. $G_1 = 0$. In this case, if $\omega_1 = H$ the option seller makes a profit without risk! This is however not an arbitrage for the seller, because it only occurs if the option holder make a sub-optimal choice.

Remark 93. In example 92, the holder of the option was indifferent between A_2 and B_1 if $\omega_1 = T$, so in this case both choices are optimal. If $\omega_1 = H$ the model says it is optimal for the holder to choose A_2, but in the real world there are often reasons why the holder would not exercise in the way the model says is optimal. For example, buyers of American options rarely exercise at the theoretically optimal time. For example, a company which runs an industrial process that needs some metal, say copper, can buy an American call option with strike K on copper to make sure that under no circumstances it will pay more than K for it. It would then often exercise such option when it actually needs to get copper for use in the production process, instead of at the theoretically optimal time of exercise. This is because the simple model normally considered do not take into account additional complication which happen in the real world; for example, if the company exercises an the optimal time, it will need to store the copper until when it will need it, which can have a cost that we ignored in our model.

3.9 Lecture 4, American Options

Let us consider in more detail the example of American options. A standard (European) option has a payoff $f(I_T)$ for some adapted functional I of S (i.e. $I_t = f(t, S_0, \dots, S_t)$). For example,

$$I_t = (S_t - K)^+, \quad I_t = S_t - \min_{u \le t} S_u, \quad I_t = (\tfrac{1}{t} \textstyle\sum_{u \le t} S_u - K)^+$$

are the payoffs functionals for the call option, the floating lookback call option, and the Asian option. Given an European option with payoff functional I, we can consider the corresponding American option, which has payoff I_τ at time τ, where the (stopping) time $\tau \le T$ is chosen by the option's buyer. In this case, I is the called *intrinsic value* of the derivative, τ the *exercise* date, and T is the expiry of the option. If we work on a model with expiry N (i.e. time index $0, 1, \dots, N$), the buyer can choose $\tau \le N$, or $\tau = \infty$. He will only choose $\tau(\omega) = t \le N$ if $I_t(\omega) \ge 0$; if $I_t(\omega) < 0$ for all t, he will choose $\tau = \infty$, which means the option does not get exercised (so the buyer gets paid $I_\infty = 0$). In particular, the American option with intrinsic value $(S_t - K)^+$ has the same value than the American option with intrinsic value $S_t - K$, since the buyer will simply choose not to get paid anything when $(S_t - K)^+ = 0$ (i.e. when $S_t - K \leq 0$).

To price American options, denote with V_n the value at time n of the American option (with intrinsic value I) which hasn't yet been exercised. At the final time, if you have not yet exercised, you can choose to either get paid I_N^+ or nothing, so the value is

$$V_N = \max(I_N, 0) = I_N^+. \tag{75}$$

At time n, if you have already exercised the option in the past, it is now worth nothing. If instead you have not yet exercised, you can choose to either get paid now the amount I_n, or to wait until time $n+1$. In the latter case you will own at time $n+1$ an American option which hasn't yet been exercised. The value of owning this is V_{n+1} at time $n+1$; its value at time n is thus

$$\mathcal{R}_n(V_{n+1}) = \mathbb{E}^{\mathbb{Q}}\left[\frac{V_{n+1}}{1+r}\middle| X_1, \dots, X_n\right],$$

see remark 87. So, V_n can be calculated for all n by backward induction from eq. (75) using the formula

$$V_n = \max\left(I_n, \mathbb{E}^{\mathbb{Q}}\left[\frac{V_{n+1}}{1+r}\middle| X_1, \dots, X_n\right]\right), \quad n = 0, \dots, N-1. \tag{76}$$

This allows to compute V_0, which is simply the value of the American option at time 0.

Remark 94. Clearly $I^+ \leq V$ (meaning $I_n^+ \leq V_n$ for all n), and the discounted value $\bar{V}$ is a supermartingale (with respect to the risk-neutral measure $\mathbb{Q}$), meaning

$$\mathbb{E}^{\mathbb{Q}}[\bar{V}_{n+1} | X_1, \dots, X_n] \leq \bar{V}_n \quad \text{for all } n.$$

Moreover, if $Y \geq I^+$ and $\overline{Y}$ a supermartingale then $Y \geq V$.

If $I_0 < V_0$ then buyer's optimal choice is not to exercise at time 0, since this leads to a payoff I_0 which is smaller than the value of exercising at some later time. If instead $I_0 = V_0$ then the buyer's optimal choice is to exercise at time 0, since waiting would lead to a payoff which is not bigger, and which may well be smaller (it will be smaller if at no future time k it happens that $I_k = V_k$). This reasoning shows that the optimal exercise time for the buyer of the American option is

$$\tau^* := \inf\{n \leq N : I_n = V_n\};$$

and in particular the buyer should not exercise the option iff $\tau^* = \infty$.

Remark 95. In the real world, often American option get exercised at times which are not optimal. To understand the reason, suppose a company needs to periodically buy large quantities of copper, and that to hedge against the risk of rising copper prices it bought an American call option for some amount of copper. The company may not house a lot of financial experts to determine what is the optimal exercise time of the derivative; it will instead simply exercise the option at the time k at which it actually needs to buy copper (if the option is in the money, i.e. if $I_k > 0$).

3.10 Lecture 5, Conditional probability and conditional expectation

To get a better understanding of how to price derivatives in a multi-period model, we need to talk about conditional probability and conditional expectation. We discuss here these topics briefly, focusing on the case of a finite σ-algebras (which is all we need when considering a finite probability space, as is done in the binomial model in finite time).

Recall that, given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a $B \in \mathcal{A}$, when $\mathbb{P}(B) > 0$ we can define the *conditional probability* $\mathbb{P}$ *given* B as the following function

$$\mathbb{P}(\cdot|B) : \mathcal{A} \to [0,1], \quad \mathbb{P}(A|B) := \mathbb{P}(A \cap B)/\mathbb{P}(B) \quad \text{for } A \in \mathcal{A}.$$

Theorem 96. *$\mathbb{P}(\cdot|B)$ is a probability on $\mathcal{A}$, and*

$$\mathbb{E}^{\mathbb{P}(\cdot|B)}[X] = \mathbb{E}^{\mathbb{P}}[X 1_B]/\mathbb{P}(B) \tag{77}$$

for any random variable $X \geq 0$, and thus for all X s.t. $\mathbb{E}^{\mathbb{P}(\cdot|B)}[|X|] = \mathbb{E}^{\mathbb{P}}[|X|1_B] < \infty$.

Proof. If $A_n \in \mathcal{A}, n \in \mathbb{N}$ are disjoint then so are the sets $(A_n \cap B)_n$ and so

$$\mathbb{P}((\cup_n A_n) \cap B) = \mathbb{P}(\cup_n (A_n \cap B)) = \sum_n \mathbb{P}(A_n \cap B),$$

and dividing times $\mathbb{P}(B)$ we get that $\mathbb{P}(\cup_n A_n|B) = \sum_n \mathbb{P}(A_n|B)$. Since $\mathbb{P}(\cdot|B) \geq 0$ and $\mathbb{P}(\Omega|B) = 1$, $\mathbb{P}(\cdot|B)$ is a probability. Equation (77) holds if X is an indicator $X = 1_A, A \in \mathcal{A}$ (by definition of $\mathbb{P}(\cdot|B)$), and so also for any linear combination of indicators $X = \sum_{k=1}^n c_k 1_{A_k}$ (by linearity), so for any rv $X \geq 0$, by taking limits[100]. Finally, if $X \in L^1(\mathbb{P}|B)$ the result follows applying eq. (77) to the positive $X^+ \geq 0$ and negative $X^- \geq 0$ parts of $X = X^+ - X^-$, by subtraction. $\square$

The interpretation of $\mathbb{P}(\cdot|B)$ is as follows. If there is some phenomenon of which we do not know for sure some characteristic $\omega \in \Omega$, but for which we do have some educated guesses, we can model our lack of knowledge using a probability $\mathbb{P}$, which describes how likely ω is to belong to a collection (a σ-algebra) $\mathcal{A}$ of sets, called events. If we now perform some experiment, whose result doesn't exactly identify ω but does inform us that $\omega \in B$, then we should update our description of how likely each event is to occur using the newly obtained information, and thus replace $\mathbb{P}$ with the conditional probability $\mathbb{P}(\cdot|B)$. Note moreover that eq. (77) shows that $\mathbb{E}^{\mathbb{P}(\cdot|B)}(X)$ is the average of X on the event B.

Suppose now that we perform an experiment, which has n possible distinct results. Its outcome outcome tells us which B_k is s.t. that $\omega \in B_k$, where $\{B_k\}_{k=1}^n =: \Pi$ is a given (finite) *partition* of Ω, i.e. a family of disjoint, non-empty sets whose union is Ω. If the outcome of the experiment is that $\omega \in B_k$, then we should, as the probability of the event A, consider not $\mathbb{P}(A)$ but rather $\mathbb{P}(A|B_k)$. In other words, the *conditional*

[100] To be rigorous, here one has to use measure theory, specifically the monotone convergence theorem.

probability of A given Π should be defined as the random variable

$$\mathbb{P}(A|\Pi) := \sum_{k=1}^{n} 1_{B_k}\mathbb{P}(A|B_k) = \begin{cases} \mathbb{P}(A|B_1) & \text{if } \omega \in B_1 \\ \dots & \text{if } \dots \\ \mathbb{P}(A|B_n) & \text{if } \omega \in B_n \ . \end{cases} \tag{78}$$

A particularly common example is obtained starting from a random variable X which only takes finitely many values $x_1, \dots, x_n$, by which we mean that

$$\mathbb{P}(\{X = x_k\}) > 0 \quad \forall k = 1, \dots, n, \quad \mathbb{P}(X \notin \cup_{k=1}^{n}\{x_k\}) = 0, \tag{79}$$

and taking $\Pi_X := (B_k)_{k=1}^n$ for $B_k := \{X = x_k\}$; in this case $\mathbb{P}(A|X) := \mathbb{P}(A|\Pi_X)$ is called the *conditional probability of A given X*. For notational convenience, improperly we will often write $\mathbb{P}(\{X = x_k\}), \mathbb{P}(A|\{X = x_k\})$ as $\mathbb{P}(X = x_k), \mathbb{P}(A|X = x_k)$.

In the above, what mattered was not X, but rather the family $\{X = x_k\}_{k=1}^n$, so if Y is another random variable which only takes the n values $y_1, \dots, y_n$, and $\{X = x_k\}_{k=1}^n = \{Y = y_k\}_{k=1}^n$, then $\mathbb{P}(A|X) = \mathbb{P}(A|Y)$, since $\Pi_X = \Pi_Y$. Just like the random variable X leads to the finite partition $\Pi_X := \{X = x_k\}_{k=1}^n$, given a finite partition $\Pi' = \{B_k\}_{k=1}^n$ of Ω one can define the random variable $X := \sum_{k=1}^n x_k 1_{B_k}$, which is such that $\Pi_X = \Pi'$, where $\{x_k\}_{k=1}^n$ are distinct points (arbitrarily chosen), normally taken in $\mathbb{R}$ (or more generally in $\mathbb{R}^m$).

Notice that the random variable $\mathbb{P}(A|X)$ is a function of X; in other words, $\mathbb{P}(A|X)$ is $\sigma(X)$-measurable. These observations suggest that one should more generally be able to define $\mathbb{P}(A|X)$ for any random variable X, and analogously $\mathbb{P}(A|\mathcal{F})$ for any σ-algebra $\mathcal{F}$, and that $\mathbb{P}(A|\mathcal{F})$ should then be a $\mathcal{F}$-measurable random variable for every $A \in \mathcal{F}$. While this can be done, we will for now restrict ourselves to the simpler case considered above, i.e. to the case where the σ-algebra $\mathcal{F} = \sigma(\Pi)$ is generated by a finite partition Π; in this case $\mathbb{P}(A|\mathcal{F})$ is defined as $\mathbb{P}(A|\Pi)$; equivalently, given X which takes finitely many values and is s.t. $\mathcal{F} = \sigma(X)$, we define[101] $\mathbb{P}(A|\mathcal{F}) := \mathbb{P}(A|X)$.

It is thus of interest to understand when a σ-algebra is generated by a finite partition. To do that, we need to be able to characterise what elements of $\mathcal{F} := \sigma(\Pi)$ belong to Π, and ideally to explicitly construct them; let us do that. Clearly, the σ-algebra generated by a finite partition Π of Ω is given by

$$\sigma(\Pi) = \{\cup_{P \in I} P | I \subseteq \Pi\}; \tag{80}$$

in particular $\sigma(\Pi)$ is finite, and the elements of Π are the smallest (non-empty) elements of $\sigma(\Pi)$, i.e. its atoms, in the following sense.

Definition 97. An *atom* of a σ-algebra $\mathcal{F}$ is a non-empty $A \in \mathcal{F}$ such that $B \subseteq A, B \in \mathcal{F}$ imply that either $B = A$ or $B = \emptyset$. The family of atoms of $\mathcal{F}$ is denoted by $\mathcal{A}(\mathcal{F})$.

[101] Of course this definition does not depend on the choice of X s.t. $\mathcal{F} = \sigma(X)$, since $\mathbb{P}(A|X)$ only depends on the family of sets $(\{X = x_k\})_k$, not on the values $(x_k)_k$.

Theorem 98. *Assume $\mathcal{F}$ is a finite σ-algebra on Ω, and denote with $A_{\mathcal{F}}(\omega)$ the intersection of all the $A \in \mathcal{F}$ which contain $\omega \in \Omega$. Then $A_{\mathcal{F}}(\omega)$ is the smallest $A \in \mathcal{F}$ which contains ω. The family $\mathcal{A}(\mathcal{F})$ of atoms of $\mathcal{F}$ is a finite partition of Ω, and*

$$\mathcal{A}(\mathcal{F}) = \{A_{\mathcal{F}}(\omega) : \omega \in \Omega\}, \tag{81}$$

so $A_{\mathcal{F}}(\omega)$ is the (only) atom of $\mathcal{F}$ containing $\omega \in \Omega$. The function $\Pi \mapsto \sigma(\Pi)$, mapping finite partitions of Ω to finite σ-algebras on Ω, is bijective and has inverse $\mathcal{F} \mapsto \mathcal{A}(\mathcal{F})$; in particular, if $\mathcal{F} = \sigma(X)$ for some random variable X which satisfies eq. (79) then $\mathcal{A}(\mathcal{F}) = \{X = x_k\}_{k=1}^n$.

Proof. Since $\mathcal{F}$ is finite, the intersection of all the $F \in \mathcal{F}$ which contain $\omega \in \Omega$ belongs to $\mathcal{F}$, contains ω, and is such the smallest possible such set (i.e. $\omega \in A_{\mathcal{F}}(\omega) \in \mathcal{F}$, and $\omega \in B \in \mathcal{F}$ implies $A_{\mathcal{F}}(\omega) \subseteq B$). Clearly $A_{\mathcal{F}}(\omega)$ is an atom, because if $B \subsetneq A_{\mathcal{F}}(\omega), B \in \mathcal{F}$ then either $\omega \in B$, in which case $B = A_{\mathcal{F}}(\omega)$ (by the minimality of $A_{\mathcal{F}}(\omega)$), or $\omega \notin B$, in which case $B = \emptyset$ (because otherwise $\omega \in A := A_{\mathcal{F}}(\omega) \setminus B \in \mathcal{F}$, contradicting the minimality of $A_{\mathcal{F}}(\omega)$). Conversely, if A is an atom, choose $\omega \in A$, then $A_{\mathcal{F}}(\omega) \subseteq A$ (since $\omega \in A \in \mathcal{F}$ implies $A_{\mathcal{F}}(\omega) \subseteq A$) and $A_{\mathcal{F}}(\omega) \supseteq A$ (since $\omega \in B := A_{\mathcal{F}}(\omega) \cap A \in \mathcal{F}$, and so $\emptyset \neq B \subseteq A$, which implies $B = A$ by definition of atom), proving eq. (81). It follows that $\{A_{\mathcal{F}}(\omega) : \omega \in \Omega\}$ is a finite partition of Ω, since all atoms are disjoint (by definition), they are finitely many (since $\mathcal{A}(\mathcal{F}) \subseteq \mathcal{F}$), and $\omega \in A_{\mathcal{F}}(\omega) \in \mathcal{F}$ imply that $\cup_{\omega \in \Omega} A_{\mathcal{F}}(\omega) = \Omega$.

Clearly if Π is a finite partition of Ω then $\{A_{\sigma(\Pi)}(\omega) : \omega \in \Omega\} = \Pi$. Conversely, we proved that if $\mathcal{F}$ is a finite σ-algebra on Ω then $\Pi := \{A_{\mathcal{F}}(\omega) : \omega \in \Omega\}$ is a finite partition of Ω and, as we will now show, $\sigma(\Pi) = \mathcal{F}$, and so $\Pi \mapsto \sigma(\Pi)$ and $\mathcal{F} \mapsto \mathcal{A}(\mathcal{F})$ are inverse of one another. Indeed eq. (80) shows that $\sigma(\Pi) \subseteq \mathcal{F}$, and that if $F \in \mathcal{F}$ then $A := \cup_{\omega \in F} A_{\mathcal{F}}(\omega) \supseteq F$ belongs to $\sigma(\Pi)$, so the conclusion follows from $A \subseteq F$, which holds since $A_{\mathcal{F}}(\omega) \subseteq F$ for all $\omega \in F$ (by minimality of $A_{\mathcal{F}}(\omega)$). □

Notice that eq. (80) and eq. (81) allow us to explicitly construct $\sigma(\Pi), \mathcal{A}(\mathcal{F})$ given $\Pi, \mathcal{F}$.

Just like using $\mathbb{P}$ one can then build the expectation $\mathbb{E}^{\mathbb{P}}(W)$ of a random variable W, using the conditional probability $\mathbb{P}(\cdot|X)$, one can consider the *conditional* $\mathbb{P}$*-expectation of* W *given* X, defined as the random variable

$$\mathbb{E}^{\mathbb{P}}[W|X](\omega) := \mathbb{E}^{\mathbb{P}(\cdot|X)(\omega)}[W].$$

In other words, $\mathbb{E}^{\mathbb{P}}[W|X]$ is the random variable which on $\{\omega : X(\omega) = x_k\}$ equals

$$\mathbb{E}^{\mathbb{P}}(W|X = x_k) := \mathbb{E}^{\mathbb{P}(\cdot|X=x_k)}(W) = \mathbb{E}^{\mathbb{P}}[W 1_{\{X=x_k\}}]/\mathbb{P}(X = x_k),$$

where the equality follows from theorem 96; recall that, by definition of expectation,

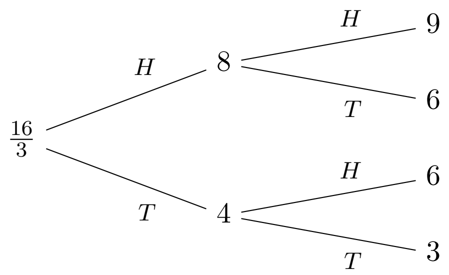
$$\mathbb{E}^{\mathbb{P}}(W|X = x_k) := \mathbb{E}^{\mathbb{P}(\cdot|X=x_k)}(W) = \textstyle\sum_i w_i \mathbb{P}(W = w_i|X = x_k) \tag{82}$$

if W takes finitely many values $\{w_i\}_i$. Thus, $\mathbb{E}(W|X)$ is the local average of W: it is the random variable which at any $\omega \in \{X = x_k\}$ equals the average of W on $\{X = x_k\}$, i.e. on the only set of the partition $\{X = x_i\}_{i=1}^n$ which contains the point ω. Analogously of course, if Π is a finite partition of Ω and $\mathcal{F}$ a finite σ-algebra on Ω, we can define $\mathbb{E}^{\mathbb{P}}[W|\Pi] := \mathbb{E}^{\mathbb{P}(\cdot|\Pi)}[W]$ and $\mathbb{E}^{\mathbb{P}}[W|\mathcal{F}] := \mathbb{E}^{\mathbb{P}(\cdot|\mathcal{F})}[W]$.

Example 99. Consider the binomial model with maturity $N = 2$, i.e. we take $\Omega = \{H, T\}^2 = \{HH, HT, TH, TT\}$ and $\mathcal{A} := \mathcal{P}(\Omega)$, and assume that

$$\mathbb{P}(HH) = \frac{1}{9}, \quad \mathbb{P}(HT) = \frac{2}{9}, \quad \mathbb{P}(TH) = \frac{1}{3}, \quad \mathbb{P}(TT) = \frac{1}{3}.$$

Assume now that the stock price S is given by the tree



and let us compute $\mathbb{E}[S_2|S_1]$: since

$$\mathbb{P}(S_2 = 6|S_1 = 8) = \tfrac{\mathbb{P}(S_1=8, S_2=6)}{\mathbb{P}(S_1=8)} = \tfrac{\mathbb{P}(HT)}{\mathbb{P}(HH,HT)} = \tfrac{\frac{2}{9}}{\frac{1}{9}+\frac{2}{9}} = \tfrac{2}{3}$$

and $\mathbb{P}(\cdot|S_1 = 8)$ is a probability, we get that

$$\mathbb{P}(S_2 = 9|S_1 = 8) = 1 - \mathbb{P}(S_2 = 6|S_1 = 8) = \frac{1}{3},$$

and so

$$\mathbb{E}(S_2|S_1 = 8) = 6 \cdot \frac{2}{3} + 9 \cdot \frac{1}{3} = 7.$$

The computation of $\mathbb{P}(S_2 = 9|S_1 = 4), \mathbb{E}(S_2|S_1 = 4)$ is analogous and gives

$$\mathbb{P}(S_2 = 6|S_1 = 4) = \cdots = \frac{1}{2}, \quad \mathbb{E}(S_2|S_1 = 4) = 6 \cdot \frac{1}{2} + 3 \cdot \frac{1}{2} = \frac{9}{2}.$$

Thus

$$\mathbb{E}(S_2|S_1) = 7 1_{\{S_1=8\}} + \frac{9}{2} 1_{\{S_1=4\}}.$$

Week 8

3.11 Lecture 1, Properties of the conditional expectation

As the RNPF makes clear, the notion of conditional expectation will be crucially important to us, so let us explore it in more detail. First, we define the conditional expectation with respect to arbitrary σ-algebras; then, we need to study its properties.

We have so far defined $\mathbb{E}[X|\mathcal{G}]$ for any finite σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ as a local average of X, i.e. $\mathbb{E}[X|\mathcal{G}]$ takes, on every atom of $\mathcal{G}$, the average value that X takes on that atom. One could then define $\mathbb{E}[X|\mathcal{G}]$, for an arbitrary σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, as the limit in L^1 of $\mathbb{E}[X|\mathcal{H}]$ as the finite sigma algebras $\mathcal{H} \subseteq \mathcal{G}$ become bigger and bigger. This approach can be made rigorous with the following definition, which is formally identical to the definition of limit as $t \to +\infty$ of a function $x(t)$ of a real variable $t \in \mathbb{R}$ (familiar to all students), as long as one replaces the distance $|x-y|$ between real numbers $x, y \in \mathbb{R}$ with the distance $||X-Y||_{L^1} := \mathbb{E}|X-Y|$ between integrable random variables $X, Y \in L^1$. Denote with $\mathbb{H}(\mathcal{G})$ the family of all finite σ-algebras $\mathcal{H} \subseteq \mathcal{G}$, ordered by inclusion (i.e. $\mathcal{H}_1 \leq \mathcal{H}_2$ if $\mathcal{H}_1 \subseteq \mathcal{H}_2$). We say that the net[102] $(\mathbb{E}[X|\mathcal{H}])_{\mathcal{H}\in\mathbb{H}(\mathcal{G})}$ converges in L^1 to $Y \in L^1$ if, for every $\epsilon > 0$, there exist $\mathcal{H}^* \in \mathbb{H}(\mathcal{G})$ such that, for every $\mathcal{H} \in \mathbb{H}(\mathcal{G}), \mathcal{H} \geq \mathcal{H}^*$ one has $||\mathbb{E}[X|\mathcal{H}] - Y||_{L^1} < \epsilon$. We can then define $\mathbb{E}[X|\mathcal{G}]$ as the L^1-limit of $(\mathbb{E}[X|\mathcal{H}])_{\mathcal{H}\in\mathbb{H}(\mathcal{G})}$. This approach has the advantage of being an intuitive way of defining $\mathbb{E}[X|\mathcal{G}]$ as a local average; the disadvantage is that it is not obvious when such limit exists. It is however possible to give a reasonably simple proof of the fact that it always does, i.e. of the following theorem.

Theorem 100. *For all $X \in L^1$ and σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ the net $(\mathbb{E}[X|\mathcal{H}])_{\mathcal{H}\in\mathbb{H}(\mathcal{G})}$ converges in L^1. Its L^1-limit is called the conditional expectation of X given $\mathcal{G}$, denoted with $\mathbb{E}[X|\mathcal{G}]$.*

The following equivalent characterisation (due to Kolmogorov) of the notion of conditional expectation is so convenient that it is normally used as a definition[103] of $\mathbb{E}[X|\mathcal{G}]$; we prefer not to do that, since we find it less intuitive.

Theorem 101. *$Z := \mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable and satisfies*

$$\mathbb{E}[ZW] = \mathbb{E}[XW] \text{ for all } \mathcal{G}\text{-measurable and bounded } W; \tag{83}$$

moreover, $Z = \mathbb{E}[X|\mathcal{G}]$ is the unique *$\mathcal{G}$-measurable random variable[104] $Z \in L^1$ which satisfies eq. (83).*

---

[102] A net is a generalisation of the notion of sequence, for which the indexing set $\mathbb{N}$ is replaced by any (partially) ordered set I which is directed upward, i.e. such that for every $i, j \in I$ there exists $k \in I$ such that $i \leq k, j \leq k$ (obviously $\mathbb{H}(\mathcal{G})$ is one such set).

[103] In which case, one has to prove that a $\mathcal{G}$-measurable random variable Z which satisfies eq. (83) exists for every $X, \mathcal{G}$.

[104] Of course, to be precise Z, as every element of $L^1(\mathbb{P})$, is an equivalence class of random variables, so the stated uniqueness is up to $\mathbb{P}$-null sets.

Just like the expectation satisfies many useful properties which allow to perform calculations with it, so does the conditional expectation. We list below some of them; they are all easily proved when the σ-algebras $\mathcal{G}, \mathcal{H}$ are finite, and passing to the limit shows that they hold for arbitrary σ-algebras. Alternatively, these properties can be easily proved via Kolmogorov's characterisation given by theorem 101. In what follows X, Z, W are random variables, all defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and $\mathcal{G} \subseteq \mathcal{A}$ is a σ-algebra, and we assume that all quantities of which we take a conditional expectation are integrable.

1. *Linearity*: $\mathbb{E}(X + Z|\mathcal{G}) = \mathbb{E}(X|\mathcal{G}) + \mathbb{E}(Z|\mathcal{G})$
2. *Independence*: if X is independent[105] of $\mathcal{G}$ then $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$. In particular any constant $c \in \mathbb{R}$ satisfies $\mathbb{E}(c|\mathcal{G}) = c$, and if $\mathcal{G}$ is the trivial σ-algebra $\{\emptyset, \Omega\}$ then any X satisfies $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}[X]$.
3. *Taking out what is know*: if X is $\mathcal{G}$-measurable then $\mathbb{E}(XZ|\mathcal{G}) = X\mathbb{E}(Z|\mathcal{G})$, and in particular[106] $\mathbb{E}(X|\mathcal{G}) = X$.
4. *Iterated conditioning (a.k.a. tower property):* If $\mathcal{H} \subseteq \mathcal{A}$ is a σ-algebra and $\mathcal{G} \subseteq \mathcal{H}$ then $\mathbb{E}(\mathbb{E}(X|\mathcal{H})|\mathcal{G}) = \mathbb{E}(X|\mathcal{G})$; in particular[107] $\mathbb{E}[\mathbb{E}(X|\mathcal{H})] = \mathbb{E}[X]$.
5. *Jensen inequality*: if $\phi : \mathbb{R} \to \mathbb{R}$ is convex then $\mathbb{E}[\phi(X)|\mathcal{G}] \geq \phi(\mathbb{E}[X|\mathcal{G}])$

To highlight the importance of the concept of conditional expectation, we mention its following characterisation which, to be precise, applies only to the conditional expectation $\mathbb{E}[\cdot|\mathcal{G}]$ restricted to the space

$$L^2 := L^2(\mathcal{A}) := L^2(\Omega, \mathcal{A}, \mathbb{P}) := \{X : \Omega \to \mathbb{R}, X \text{ is } \mathcal{A}\text{-measurable}, \mathbb{E}[X^2] < \infty\}$$

of square integrable random variables. The best approximation of X with a constant is $\mathbb{E}[X]$, for example[108] in the sense that, if $\mathbb{E}[X^2] < \infty$, then $\mathbb{E}[X]$ is the unique constant c which minimises $\mathbb{E}[(X - c)^2]$ across all $c \in \mathbb{R}$. Analogously, the following theorem characterises the conditional expectation $\mathbb{E}[X|\mathcal{G}]$ as the best approximation of X which can be achieved using the information given by $\mathcal{G}$.

Theorem 102. *Assume $X \in L^2(\mathcal{A})$. Then $\mathbb{E}[X|\mathcal{G}]$ is the unique C minimiser of $\mathbb{E}[(X - C)^2]$ across $C \in L^2(\mathcal{A})$. Equivalently $\mathbb{E}[X|\mathcal{G}]$ is the unique $C \in L^2(\mathcal{A})$ such that*

$$\mathbb{E}[(X - C)W] = 0 \text{ for all } W \in L^2(\mathcal{G}). \tag{84}$$

[105] We discuss independence in section 3.15.
[106] Take $Z = 1$ and use the independence property $\mathbb{E}(1|\mathcal{G}) = 1$.
[107] Take $\mathcal{G}$ trivial and use the independence property.
[108] In fact, instead the function $g(x, c) = (x - c)^2$ one could choose any g in a large family of 'loss' functions, and for each of them the unique minimiser of $\mathbb{E}g(X, c)$ over $c \in \mathbb{R}$ is $c = \mathbb{E}[X]$.

Notice that theorem 102 admits the following geometrical interpretation: $\mathbb{E}[\cdot|\mathcal{G}]$ is the projection[109] of the Hilbert[110] space $H := L^2(\mathcal{A})$ onto its closed subspace

$$V := L^2(\mathcal{G}) := L^2(\Omega, \mathcal{G}, \mathbb{P}) := \{X \in L^2 : X \text{ is } \mathcal{G}\text{-measurable}\}.$$

Moreover, the characterisation eq. (84) is very similar to theorem 101.

Lemma 103. *The set $L^2 = L^2(\mathbb{P})$ of random variables X s.t. $\mathbb{E}X^2 < \infty$ is a vector space, and if $X, Y \in L^2(\mathbb{P})$ then $XY \in L^1(\mathbb{P})$.*

Proof. If $\mathbb{E}X^2 < \infty$ and $c \in \mathbb{R}$ trivially $\mathbb{E}(cX)^2 = c^2\mathbb{E}X^2 < \infty$. The trivial inequalities $(x \pm y)^2 \geq 0$ imply $2|xy| \leq x^2 + y^2$, which shows that $X, Y \in L^2(\mathbb{P})$ imply $XY \in L^1(\mathbb{P})$ and so $X + Y \in L^2(\mathbb{P})$. □

Proof of theorem 102. It follows from Jensen inequality that $W := \mathbb{E}[X|\mathcal{G}]$ is in L^2. Since $X \in L^2$, it is enough[111] to prove that $\mathbb{E}[(X-C)^2] \geq \mathbb{E}[(X-W)^2]$ for $\mathbb{E}[C^2] < \infty$; in this case, because of lemma 103 all the random variables which we are about to consider are in $L^1(\mathbb{P})$. So, to conclude assume $\mathbb{E}[C^2] < \infty$ and notice that

$$\mathbb{E}(X-C)^2 = \mathbb{E}[((X-W)+(W-C))^2] = \mathbb{E}(X-W)^2 + \mathbb{E}(W-C)^2 + 2\mathbb{E}(X-W)(W-C),$$

which is bigger than $\mathbb{E}[(X-W)^2]$ because $(W-C)^2 \geq 0$, and

$$\mathbb{E}[(X-W)(W-C)|\mathcal{G}] = (W-C)\mathbb{E}[X-W|\mathcal{G}] = (W-C)(W-W) = 0,$$

which implies $\mathbb{E}[(X-W)(W-C)] = 0$. □

Notice that the interpretation of conditional expectation afforded by theorem 102 makes the properties of Independence, Taking out what is know, and Iterated conditioning, somewhat intuitive.

3.12 Lecture 2, The RNPF in the multi-period binomial model

As the above example 86 indicates, to price in the multi-period binomial model, for $\omega \in \{H, T\}^N, n \leq N$ we set $\omega(n) := (\omega_1, \ldots, \omega_n)$ and compute the up and down factors

$$U_n(\omega) := \frac{S_{n+1}((\omega(n), H))}{S_n(\omega(n))}, \quad D_n(\omega) := \frac{S_{n+1}((\omega(n), T))}{S_n(\omega(n))},$$

and define the risk-neutral transition-probabilities $\tilde{P}_n$ and $\tilde{Q}_n := 1 - \tilde{P}_n$ by asking that

$$\overline{S}_n(\omega(n)) = \tilde{P}_n(\omega(n))\overline{S}_{n+1}((\omega(n), H)) + \tilde{Q}_n(\omega(n))\overline{S}_{n+1}((\omega(n), T)); \qquad (85)$$

[109] If H is Hilbert with dot product $\langle a, b\rangle_H$ and norm $\|h\| := \sqrt{\langle h, h\rangle_H}$, and $V \subseteq H$ is a closed subspace thereof, the projection of $h \in H$ onto V is defined as the unique minimiser of $\|h - v\|_H$ over $v \in V$, or equivalently as the $v \in V$ such that $\langle h, v - w\rangle_H = 0$ for all $w \in V$, and such projection always exists unique.

[110] $L^2(\mathcal{A})$ with the dot product $\langle X, Y\rangle := \mathbb{E}[XY]$ is a Hilbert space.

[111] Because if $\mathbb{E}[C^2] = \infty$ then by lemma 103 $\mathbb{E}[(X-C)^2 = \infty$, and so $\mathbb{E}[(X-C)^2] = \infty > \mathbb{E}[(X-W)^2]$.

solving for $\tilde{P}_n$ gives

$$\tilde{P}_n(\omega) = \tilde{P}_n(\omega(n)) := \frac{(1+R_n) - D_n}{U_n - D_n}(\omega(n)) \quad n = 0, \dots, N-1, \tag{86}$$

$$\tilde{Q}_n(\omega) = \tilde{Q}_n(\omega(n)) := \frac{U_n - (1+R_n)}{U_n - D_n}(\omega(n)) \quad n = 0, \dots, N-1. \tag{87}$$

Then, compute $V_n = V_n^{x,G}$ by backward induction using the same formula as for $\overline{S}$:

$$\overline{V}_n(\omega(n)) = \tilde{P}_n(\omega(n))\overline{V}_{n+1}((\omega(n), H)) + \tilde{Q}_n(\omega(n))\overline{V}_{n+1}((\omega(n), T)). \tag{88}$$

It turns out we can then express the above formula in a much neater way, as follows; but first, let us describe how to use a convenient notation.

Remark 104. While to be formal we should always write e.g.

$$\mathbb{P}(\{Z = a\}|\{Y \leq b\}), \text{ and } \mathbb{Q}(\{\omega' \in \Omega : X_3(\omega') = \omega_3\}|\{\omega' \in \Omega : X_1(\omega') = \omega_1, X_2(\omega') = \omega_2\}),$$

we will often improperly and more simply write $\mathbb{P}(Z = a|Y \leq b)$ and $\mathbb{Q}(\omega_3|\omega_1\omega_2)$, so that if e.g. $\omega_1 = H, \omega_2 = T, \omega_3 = T$ then $\mathbb{Q}(\omega_3|\omega_1\omega_2) = \mathbb{Q}(T|HT)$ means

$$\mathbb{Q}(\{\omega' \in \{H,T\}^N : X_3(\omega') = T\}|\{\omega' \in \{H,T\}^N : X_1(\omega') = H, X_2(\omega') = T\})$$

or equivalently

$$\mathbb{Q}(\{\omega' \in \{H,T\}^N : \omega'_3 = T\}|\{\omega' \in \{H,T\}^N : \omega'_1 = H, \omega'_2 = T\}).$$

Recall that eq. (88) came from considering the one-period binomial model $(B_t, S_t)_{t=n,n+1}$ under the assumption that that the result of the first n coin tosses was $\omega(n)$, and $\tilde{P}_n(\omega(n))$ could then be interpreted as the risk-neutral probability of Heads. Thus, we want to consider a probability $\mathbb{Q}$ on $\Omega = \{H,T\}^N$ s.t.

$$\mathbb{Q}(H|\{X(n) = \omega(n)\}) = \tilde{P}_n(\omega(n)) \quad n = 0, \dots, N-1, \omega \in \Omega, \tag{89}$$

where as usual

$$\{X(n) = \omega(n)\} = \{\omega' : X(n)(\omega') = \omega(n)\} = \{\omega' : (X_1, \dots, X_n)(\omega') = (\omega_1, \dots, \omega_n)\}$$

(so e.g. $\mathbb{Q}(H|\{X_1 = H, X_2 = T\}) = \tilde{P}_2(HT)$). Then eq. (88) can be interpreted as saying that $\overline{V}_n(\omega(n))$ is given by the expectation of $\overline{V}_{n+1}$ with respect to the probability $\mathbb{Q}$ conditioned on the event $\{X(n) = \omega(n)\}$, i.e.

$$\overline{V}_n(\omega(n)) = \mathbb{E}^{\mathbb{Q}|\{X(n)=\omega(n)\}}[\overline{V}_{n+1}]. \tag{90}$$

Morever, the RHS (*resp. LHS*) of eq. (90) is, by definition, the constant value that the conditional expectation $\mathbb{E}^{\mathbb{Q}}[\overline{V}_{n+1}|X(n)] = \mathbb{E}^{\mathbb{Q}}[\overline{V}_{n+1}|\mathcal{F}_n]$ (*resp. that* $\overline{V}_n$) takes on the set $\{X(n) = \omega(n)\}$. Thus we get the RNPF (Risk-Neutral Pricing Formula)

$$\overline{V}_n = \mathbb{E}^{\mathbb{Q}}[\overline{V}_{n+1}|\mathcal{F}_n], \tag{91}$$

which has a very pleasant form, and extends to the multi-period setting the formula $\overline{V}_0 = \mathbb{E}^{\mathbb{Q}}[\overline{V}_1]$ which we used in one-period models.

That in the binomial model a $\mathbb{Q}$ s.t. eq. (89) holds exists, and is unique, follows from the following lemma when applied to the $\tilde{P}$ given by eq. (86).

Lemma 105. *The map $\mathbb{Q} \mapsto \tilde{P}$ given by Equation (89) is a bijection between*

1. *probabilities $\mathbb{Q}$ on $(\Omega, \mathcal{A})$*
2. *$\mathcal{F}$-adapted processes $\tilde{P} = (\tilde{P}_n)_{n \leq N}$ on $(\Omega, \mathcal{A})$ with values in $[0, 1]$*

Moreover $\mathbb{Q} \sim \mathbb{P} \iff 0 < \tilde{P} < 1$.

Proof of lemma 105. Clearly eq. (89) shows how to build $\tilde{P}$ given $\mathbb{Q}$; such $\tilde{P}$ is adapted since $\tilde{P}_n$ only depends on $\omega(n)$. To build $\mathbb{Q}$ from $\tilde{P}$, let us just consider the case $N = 2$: the idea is the same[112], but the notation is less heavy. We will write $\tilde{Q}$ for $1 - \tilde{P}$. By definition of conditioning, any probability $\mathbb{Q}$ on $\{H, T\}^2$ needs to satisfy $\mathbb{Q}(\omega_1\omega_2) = \mathbb{Q}(\omega_1)\mathbb{Q}(\omega_2|\omega_1)$, and so eq. (89) determine $\mathbb{Q}$ as follows:

$$\mathbb{Q}(HT) = \mathbb{Q}(H)\mathbb{Q}(T|H) = \mathbb{Q}(X_1 = H)\mathbb{Q}(X_2 = T|X_1 = H) = \tilde{P}_0\tilde{Q}_1(H),$$

and analogously we find

$$\mathbb{Q}(HH) = \tilde{P}_0\tilde{P}_1(H), \quad \mathbb{Q}(TT) = \tilde{Q}_0\tilde{Q}_1(T), \quad \mathbb{Q}(TH) = \tilde{Q}_0\tilde{P}_1(T).$$

Notice that such $\mathbb{Q}$ is a probability equivalent to $\mathbb{P}$, since $\tilde{Q} = 1 - \tilde{P}$ implies

$$\mathbb{Q}(HH) + \mathbb{Q}(HT) = \tilde{P}_0, \qquad \mathbb{Q}(TH) + \mathbb{Q}(TT) = 1 - \tilde{P}_0$$

and so $\mathbb{Q}(\Omega) = 1$, and $\mathbb{Q} > 0$ is equivalent to $0 < \tilde{P} < 1$, since the product of positive (i.e. ≥ 0) numbers is positive, and it equals 0 iff one of them equals 0. $\square$

The

$$\tilde{P}_n := \mathbb{Q}(X_{n+1} = H|\mathcal{F}_n), \quad \tilde{Q}_n := 1 - \tilde{P}_n = \mathbb{Q}(X_{n+1} = T|\mathcal{F}_n), n \leq N - 1$$

are called *transition probabilities* corresponding to the probability $\mathbb{Q}$, since $\tilde{P}_n, 1 - \tilde{P}_n$ describe the probabilities relative to the transition between time n and time $n+1$. More generally, one could consider a model where X_{n+1} takes finitely many values $\{x^k_{n+1}\}^d_{k=1}$, and consider the transition probabilities $\mathbb{Q}(X_{n+1} = x^k_n|\mathcal{F}_n), 1 \leq k \leq d$.

[112] E.g. if $N = 4$ and $\omega(n) = (H, T, T, H)$ we have

$$\mathbb{Q}(HTTH) = \mathbb{Q}(H)\mathbb{Q}(T|H)\mathbb{Q}(T|HT)\mathbb{Q}(H|HTT) = \tilde{P}_0\tilde{Q}(H)_1\tilde{Q}_2(HT)\tilde{P}_3(HTT).$$

Remark 106. Often, but not always, it is more convenient to use the local[113] description of a probability given by the transition probability $\mathbb{Q}(\cdot|\mathcal{F}_n)$, rather than the global[114] description $\mathbb{Q}$. Too mention an example, use eq. (91) and the tower property[115] of the conditional expectation to get that

$$\overline{V}_k = \mathbb{E}^{\mathbb{Q}}[\overline{V}_n|\mathcal{F}_k] \text{ for any } 0 \leq k \leq n \leq N, \tag{92}$$

and in particular

$$\overline{V}_0 = \mathbb{E}^{\mathbb{Q}}[\overline{V}_N] \tag{93}$$

with which we could compute the initial value without having to compute all the intermediate prices by backward induction. However, unless $\tilde{P}$ is constant, it is in general harder to compute $\mathbb{Q}$ from $\tilde{P}$ and then use eq. (93) to compute $\overline{V}_0$, than it is to compute $\overline{V}_0$ directly from $\tilde{P}$ using backward induction.

3.13 Lecture 3, The FTAP in the multi-period setting

As $\tilde{P}$ is determined by eq. (85), which is eq. (88) with $V = S$, eq. (91) states that $\mathbb{Q}$ satisfies

$$\overline{S}_n = \mathbb{E}^{\mathbb{Q}}[\overline{S}_{n+1}|\mathcal{F}_n]. \tag{94}$$

This leads us to the following definitions, valid in the setting of an arbitrary filtered probability space $(\Omega, \mathcal{A}, \mathcal{F}, \mathbb{P})$.

Definition 107.

1. An adapted process $Y = (Y_t)_{t\in\mathbb{T}}$ is a *martingale* if $Y_t \in L^1(\mathbb{P})$ and $\mathbb{E}[Y_t|\mathcal{F}_s] = Y_s$ for each $s \leq t, s, t \in \mathbb{T}$.
2. A proba $\mathbb{Q}$ on $\mathcal{F}$ is a *Martingale Measure* (for $\overline{S} := S/B$) if $\overline{S}$ is a martingale on $(\Omega, \mathcal{A}, \mathcal{F}, \mathbb{Q})$ (we say simply: if $\overline{S}$ is a $\mathbb{Q}$-martingale). Such $\mathbb{Q}$ is an EMM (Equivalent MM) if $\mathbb{Q} \sim \mathbb{P}$. The set of EMM is denoted by $\mathcal{M}(\overline{S})$.

As in the one-period case one can get:

Theorem 108 (1st FTAP)**.** *A multi-period market $(B_t, S_t)_{t\in\mathbb{T}}$ is arbitrage free* $\iff$ $\mathcal{M}(\overline{S}) \neq \emptyset$.

Proof. ($\Longrightarrow$) In the very special and simple case of the binomial model, this implication follows from the fact that if there is no arbitrage then $D < 1 + T < U$ (by theorem 85), and so eq. (86) gives $0 < \tilde{P} < 1$, so lemma 105 provides us with a $\mathbb{Q} \in \mathcal{M}(\overline{S})$. For finite Ω, one could prove the implication using LP, but we skip this. For general Ω the proof is very difficult: the theorem was only rigorously proved in 1990, about 30 years after the

[113] Local in the sense that it clarifies what happens when moving locally, i.e. between time n and time $n + 1$.

[114] Global since it assigns probabilities to *whole* trajectory $(\omega_1, \ldots, \omega_n)$.

[115] i.e. the fact that $X_i := \mathbb{E}^{\mathbb{Q}}[Y|\mathcal{F}_i]$ satisfies $X_k = \mathbb{E}^{\mathbb{Q}}[X_n|\mathcal{F}_k]$ for all $0 \leq k \leq n$; we will discuss the properties of the conditional expectation later on.

discovery of Black and Scholes formula! We refer the interested reader to [**?**, Theorem 6.1.1]. ($\Longleftarrow$) Let us assume by contradiction that G is an arbitrage, so $\overline{V}_N^{0,G} \geq 0$ $\mathbb{P}$ a.s. and so $\mathbb{Q}$ a.s., and $\{\overline{V}_N^{0,G} > 0\}$ is not a null set under $\mathbb{P}$ and thus also under $\mathbb{Q}$; it follows that $\mathbb{E}^{\mathbb{Q}}(\overline{V}_N^{0,G}) > 0$. This contradicts the fact that, since $\mathbb{Q}$ is a MM for (B, S),

$$\mathbb{E}^{\mathbb{Q}}(\overline{V}_N^{0,G}) = \sum_{n=0}^{N-1} \mathbb{E}^{\mathbb{Q}}(G_n \cdot (\overline{S}_{n+1} - \overline{S}_n))$$

equals zero, since it is the sum over n of the expectation of

$$\mathbb{E}^{\mathbb{Q}}(G_n \cdot (\overline{S}_{n+1} - \overline{S}_n)|\mathcal{F}_n) = G_n \cdot \mathbb{E}^{\mathbb{Q}}(\overline{S}_{n+1} - \overline{S}_n|\mathcal{F}_n) = G_n \cdot 0 = 0.$$

□

As in the one-period case, from the FTAP we could derive several corollaries, e.g.

Corollary 109. *Let $(B_t, S_t)_{t\in\mathbb{T}}$ be a multi-period market free of arbitrage. Then (B, S) is complete $\iff$ the EMM is unique (i.e. $\mathcal{M}(\overline{S})$ is a singleton).*

We skip the details, as in the multi-period setting we only consider in detail the binomial model, for which pricing is simple because the model is complete.

3.14 Lecture 4, Permutation-invariant processes and recombinant trees

Let us assume that we work in the simplest and nicest of settings: the binomial model. To find the AFP $Y := (Y_n)_{n\leq N}$ of a derivative with payoff Y_N we work by backward induction, setting $\overline{V}_N := \overline{V}_N^{x,G} := \overline{Y}_N$ and using the RNPF eq. (91) to compute $\overline{V}_n = \mathbb{E}^{\mathbb{Q}}(\overline{V}_{n+1}|\mathcal{F}_n)$ and G_n for each $0 \leq n \leq N-1$. This leads to a major problem: the amount of computation required to compute $\overline{V}_n = \mathbb{E}^{\mathbb{Q}}(\overline{V}_{n+1}|\mathcal{F}_n)$ is proportional to the # of paths $(\omega_1, \cdots, \omega_n) = \#\{H, T\}^n = 2^n$, which growths exponentially with n. As the binomial models used in practice have $N \geq 100$, and $2^{100} \sim 10^{30}$, computing P applying plainly the RNPF as done above is impossible even on a powerful computer. What can we do?

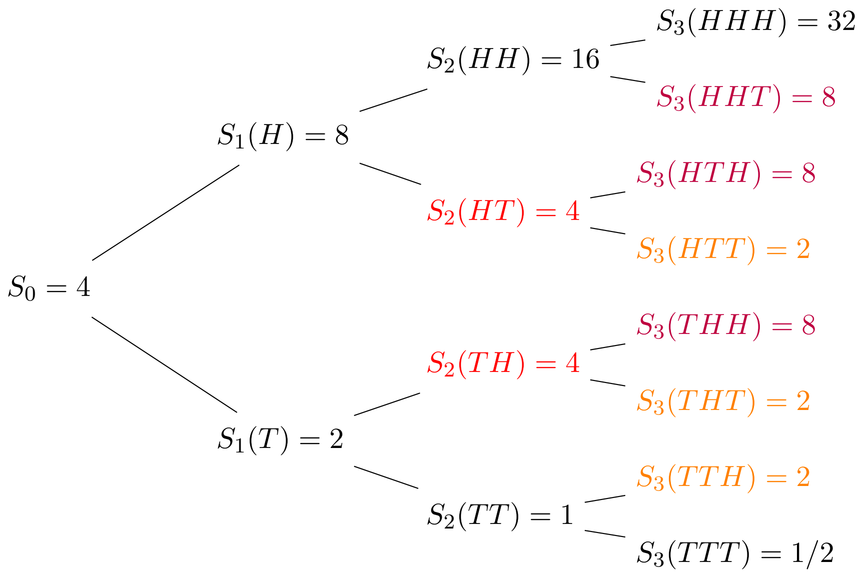
The first step in the solution is to consider for (B, S) only those adapted processes W such that, for each n, W_n takes the same value at the point $(\omega_1, \ldots, \omega_n)$ as at the point $(\sigma(\omega_1), \ldots, \sigma(\omega_n))$ where σ is any permutation of $\Omega_n := \{H, T\}^n$ (i.e. a bijection from Ω_N to Ω_N); we[116] will call such W *permutation-invariant.* If W is permutation-invariant, W_n takes at most $n+1$ possible values, as it is a function only of the number $k = 0, \cdots, n$ of coin tosses which result in Heads. As the number of values of W_n grows *linearly* in n, instead of exponentially as it did for arbitrary (B, S), we might not encounter the same computational problems when pricing an option with arbitrary underlying (B, S).

[116] We just made this terminology, since we are not aware of any commonly accepted expression for such W.

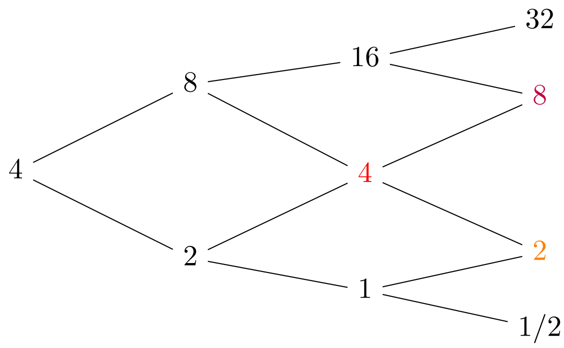
The typical example of a permutation-invariant process $(\overline{B}, \overline{S})$ is obtained by considering only underlying S with $U, D, R = u, d, r$ constants (in both t and ω), in which case we get that $\overline{B} = 1$ and $\overline{S}_n$ takes only the values

$$\{\overline{S}_n(\omega) : \omega \in \Omega_N\} = \left\{ \frac{S_0 u^k d^{n-k}}{(1+r)^n} : \ k = 0, \cdots, n \right\}.$$

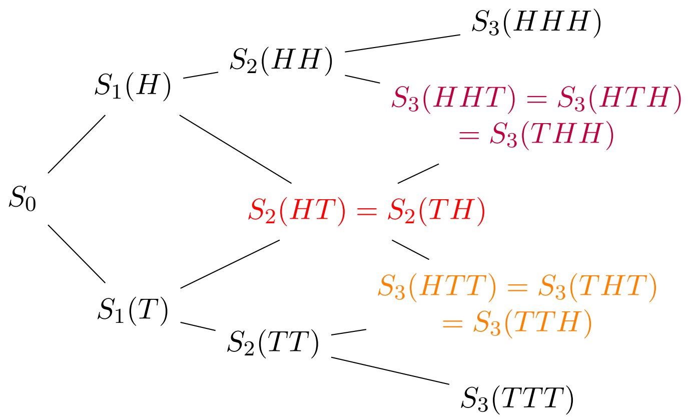
Consider for example the case of one underlying S and parameters $S_0 = 4, u = 2 = 1/d, r = 0$; then $B = 1$ and S is represented by the tree



Since S is permutation-invariant, its tree involves some repetition, which we highlighted using colors. So we could choose to more simply represent S with the following tree



In this latter tree we have collapsed repeating nodes into one; more precisely, the nodes at time n which correspond to multiple sequences of n coin tosses resulting in the same number of Heads have been collapsed into one node. More generally, if S is permutation-invariant and $N = 3$, we could represent it with the tree



The one-to-one correspondence which maps adapted processes on Ω_N to binary trees, maps permutation-invariant processes one-to-one to *recombining* (a.k.a. *recombinant*) binomial trees, i.e. trees whose branches merge back together. Recombinant trees represent permutation-invariant processes in a conveniently compact, efficient way, which cuts down on the unnecessary repetition. Analogously, when we store in computer memory the values of permutation-invariant processes, we should do it without the unnecessary repetition.

Suppose (B, S) is permutation-invariant. If the value $\overline{Y}_N$ of a derivative at time N depended just on (B_N, S_N), i.e. $\overline{Y}_N = f_N(B_N, S_N)$, then we would only need to keep track of the $N+1$ values of (B_N, S_N) to compute $\overline{Y}_N$. For a general derivative we would have $\overline{Y}_N = f_N((B_k, S_k)_{k \le N})$, we would still only need to keep track of $\sum_{k=0}^{n}(k+1) = \frac{1}{2}(N+1)(N+2) \sim N^2$ numerical values to express the discounted payoff $\overline{Y}_N$; this is fine, since N^2 does not grow too fast in N.

However for $n < N$ in general the random variable $\overline{Y}_n = \mathbb{E}^{\mathbb{Q}}(\overline{Y}_{n+1}|\mathcal{F}_n)$ will be simply $\mathcal{F}_n$-measurable, i.e. function of the coin tosses $X(n) = (X_1, \dots, X_n)$. Since $X(n)$ takes $\#\{H, T\}^n = 2^n$ possible values, to be able to actually compute $\overline{Y}_n$ we need to prove that it is a function only of W_n, , i.e. $\overline{Y}_n = f_n(W_n)$, where W_n only takes 'few' values, *for every* n. We already know that this is true at time $n = N$ (taking e.g. $W_N = (B, S_1, \dots, S_N)$), and we need to find a way to conclude that this also holds at previous times. This can be done for most derivatives using the concept of Markov process, which we will soon introduce. To do so, we first need to talk about independence, and to study in more detail the concept of conditional expectation.

3.15 Lecture 5, Independence

The following definition is of central importance in probability theory.

Definition 110. Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, we say two events $A, B \in \mathcal{A}$ are *independent* (under $\mathbb{P}$; one can also say $\mathbb{P}$-independent) if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

To shine some light on the previous definition, consider that the equality $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ is trivially satisfied if $\mathbb{P}(B) = 0$ (because $A \cap B \subseteq B$ implies $\mathbb{P}(A \cap B) \le \mathbb{P}(B)$), and if instead $\mathbb{P}(B) = 0$ it satisfied iff $\mathbb{P}(A|B) = \mathbb{P}(A)$, which means that the knowledge of whether B has occurred or does not change the probability of A, corresponding with the intuitive meaning of independence.

Given $A, B \in \mathcal{A}$, notice that if

$$\mathbb{P}(C \cap B) = \mathbb{P}(C)\mathbb{P}(B) \tag{95}$$

holds for $C = A$ then

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A^c)\mathbb{P}(B),$$

i.e. eq. (95) also holds for $C = A^c$; by addition it holds if $C = \Omega$, and it trivially holds for $C = \emptyset$. Thus, eq. (95) holds for $C = A$ iff it holds for all $C \in \sigma(A)$. It turns out however that, given events $A, B, C \in \mathcal{A}$, the identity

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$$

does *not* imply the identity

$$\mathbb{P}(A^c \cap B \cap C) = \mathbb{P}(A^c)\mathbb{P}(B)\mathbb{P}(C),$$

and that if we want to generalise the notion of independence to families containing more than two sets, and obtain a useful notion, we need to ask that both the above identities hold. This leads us to the following definition.

Definition 111. Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, we say the (finitely many) events $F_i \in \mathcal{A}, i \in J = \{i_j\}_{j=1}^n$ are $\mathbb{P}$-*independent* if we have

$$\mathbb{P}\left(G_{i_1} \cap ... \cap G_{i_n}\right) = \prod_{j=1}^n \mathbb{P}\left(G_{i_j}\right), \tag{96}$$

for any $G_{i_j} \in \{F_{i_j}, F_{i_j}^c\}, j = 1, \ldots, n$.

Remark 112. Trivially, if eq. (96) holds for any $G_{i_j} \in \{F_{i_j}, F_{i_j}^c\}$, then it holds for any

$$G_{i_j} \in \sigma(F_{i_j}) = \{\Omega, F_{i_j}, F_{i_j}^c, \emptyset\}.$$

Thus, $\{G_i\}_{i \in J}$ are independent if and only if $\{G_i\}_{i \in K}$ are independent for every $K \subseteq J$, since if we apply eq. (96) with $G_{i_j} = \Omega \in \sigma(F_{i_j})$ for every $i_j \in J \setminus K$, we get that

$$\prod_{i \in K} P\left(G_i\right) = \prod_{i \in K} P\left(G_i\right) \prod_{j \in J \setminus K} P(\Omega) = P\left(\left(\cap_{i \in K} G_i\right) \cap \left(\cap_{j \in J \setminus K} \Omega\right)\right) = P\left(\cap_{i \in K} G_i\right)$$

holds for any $G_j \in \{F_j, F_j^c\}, j \in K$. This suggests the following definitions.

Definition 113. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, an *arbitrary* collection of events $G_i \in \mathcal{A}, i \in I$ are independent if $\{G_i\}_{i \in J}$ are independent for every finite $J \subseteq I$. An arbitrary collection $\{\mathcal{G}_i\}_{i \in I}$ of sub-σ-algebras of $\mathcal{A}$ are *independent* if, for every choice of sets $G_i \in \mathcal{G}_i, i \in I$, the sets $\{G_i\}_{i \in I}$ are independent. An arbitrary collection of random vectors $X_i : \Omega \to \mathbb{R}^{k_i}, i \in I$ are *independent* if the σ-algebras $\{\sigma\left(X_i\right)\}_{i \in I}$ are independent.

Notice that, because of remark 112, if $X_i = 1_{A_i}, i \in I$, then $\{X_i\}_{i\in I}$ are independent iff $\{A_i\}_{i\in I}$ are independent (as it should be, given that we always identify a set with its indicator function in probability theory, and more generally in measure theory).

Theorem 114. *Given random vectors $X_j : \Omega \to \mathbb{R}^{k_j}, j = 1, \dots, n$, the following are equivalent:*

1. *$(X_j)_j$ are independent*
2. $$\mathbb{E}\left[\prod_{k=1}^{n} f_j\left(X_j\right)\right] = \prod_{k=1}^{n} \mathbb{E}\left[f_j\left(X_j\right)\right] \tag{97}$$

   holds for any bounded and Borel functions $f_j : \mathbb{R}^{k_j} \to \mathbb{R}, j = 1, ..., n$.
3. *eq. (97) holds for any functions $f_j : \mathbb{R}^{k_j} \to \mathbb{C}, j = 1, ..., n$ of the form $f_j(x) = \exp(i\ t_j \cdot x), t_j \in \mathbb{R}^{k_j}$.*

Sketch of the proof. By definition, $(X_j)_{j=1}^n$ are independent iff eq. (97) holds whenever the f_j's are indicator functions. In this case by linearity it also holds whenever each f_j is a linear combination of indicator functions, and so by taking limits it is possible to show that it holds for any Borel bounded f_j. Analogously, if eq. (97) holds for any bounded Borel function f_j of the form $f_j(x) = \exp(i\ t_j \cdot x)$, then taking limits of linear combinations of such functions we see that it holds for every bounded continuous function f_j, and thus (taking limits again) for every bounded Borel function f_j. □

Remark 115. With a similar proof, if more is known about the law of $X_j, j \le n$, then we can also choose other families of functions which satisfy eq. (97). For example:

1. if X_j has values in $\mathbb{R}^+$ for each j and eq. (97) holds for every f_j of the form $f_j(x) = \exp(t_j x)$, $t_j \in \mathbb{R}$, or
2. if each X_j has values in a bounded set B_j (meaning that $\mathbb{P}(X_j \notin B_j) = 0$) and eq. (97) holds for every f_j which is a polynomial,

then the $(X_j)_j$ are independent.

Remark 116. Using a little bit of measure theory, it is possible to show that the following are equivalent, whether $\mathbb{T} = \{0, 1, \dots, N\}$ or $\mathbb{T} = \mathbb{N}$:

1. $(X_i)_{i\in\mathbb{T}}$ are independent
2. For every $J, K \subseteq \mathbb{T}$ s.t. $J \cap K = \emptyset$ the two random vectors $Y := (X_i)_{i\in J}$ and $Z := (X_i)_{i\in K}$ are independent
3. For every $k \in \mathbb{T} \setminus \{0\}$, the two random vectors $X(k-1) := (X_i)_{i=0}^{k-1}$ and X_k are independent

We can characterise independence making use of the concept of conditional probability, as follows.

Theorem 117. *Given two σ-algebras $\mathcal{B}, \mathcal{C} \subseteq \mathcal{A}$, the following are equivalent:*

1. *$\mathcal{B}, \mathcal{C}$ are $\mathbb{P}$-independent,*
2. *for all $B \in \mathcal{B}$, the random variable $\mathbb{P}(B|\mathcal{C})$ is constant,*
3. *for every $\mathcal{B}$-measurable $Y \in L^1(\mathbb{P})$, the random variable $\mathbb{E}(Y|\mathcal{C})$ is constant,*

and in this case $\mathbb{P}(B|\mathcal{C}) = \mathbb{P}(B)$, and $\mathbb{E}[Y|\mathcal{C}] = \mathbb{E}[Y]$.

Proof. Trivially item 3 implies item 2, and so does item 1. To conclude, let us prove that item 2 implies item 1, and show how this implies $\mathbb{E}[Y|\mathcal{C}] = \mathbb{E}[Y]$ (and thus item 3 holds and $\mathbb{P}(B|\mathcal{C}) = \mathbb{P}(B)$). We only prove this for finite $\mathcal{C}$ and[117] s.t. every atom C of $\mathcal{C}$ has probability $\mathbb{P}(C) > 0$. In this case $\mathbb{P}(B|\mathcal{C}) = c$ means that $\mathbb{P}(B|C) = c$ holds for every atom C of $\mathcal{C}$, i.e. $\mathbb{P}(B \cap C) = c\mathbb{P}(C)$ holds for every atom C of $\mathcal{C}$, and thus (since the atoms are disjoint) it holds whenever C is a finite union of atoms, i.e. for every $C \in \mathcal{C}$. Taking $C = \Omega$ implies $c = \mathbb{P}(B)$. Thus $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$ holds for every $C \in \mathcal{C}$, i.e. $\mathcal{B}, \mathcal{C} \subseteq \mathcal{A}$ are $\mathbb{P}$-independent. This shows that, if Y is an indicator function of a set in $\mathcal{B}$, then $\mathbb{E}(Y|\mathcal{C}) = \mathbb{E}(Y)$. By linearity this is true if Y is a linear combination of such indicators, and by taking limits this holds for every integrable and $\mathcal{B}$-measurable Y. □

Theorem 118. *If two random variables X, Y take finitely many values $\{x_i\}_{i=1}^n, \{y_j\}_{j=1}^m$, then X, Y are independent iff, for each i, j, the two sets $\{X = x_i\}, \{Y = y_j\}$ are independent.*

Proof. It follows from theorem 117 that X, Y are independent iff $\mathbb{P}(B|\sigma(Y)) = \mathbb{P}(B)$ for all $B \in \sigma(X)$, or equivalently[118] for all atoms B of $\sigma(X)$, i.e. iff $\mathbb{P}(\{X = x_i\}|\sigma(Y)) = \mathbb{P}(\{X = x_i\})$ for each i, i.e. iff $\mathbb{P}(\{X = x_i\}|\{Y = y_j\}) = \mathbb{P}(\{X = x_i\})$ for each i. □

Remark 119. Notice that, if $\mathcal{B}$ is finite, if $\mathbb{P}(B|\mathcal{C}) = \mathbb{P}(B)$ holds for every atom B of $\mathcal{B}$ then (since atoms are disjoint) it holds for every finite union of atoms, i.e. for every $B \in \mathcal{B}$; moreover if $\mathbb{P}(B|\mathcal{C}) = \mathbb{P}(B)$ then

$$\mathbb{P}(B^c|\mathcal{C}) = 1 - \mathbb{P}(B|\mathcal{C}) = 1 - \mathbb{P}(B) = \mathbb{P}(B^c).$$

Thus, a random variable X which only takes two values x_1, x_2 is independent of a σ-algebra $\mathcal{G}$ if and only if[119] $\mathbb{P}(X = x_1|\mathcal{G})$ is constant.

Remark 120. Combining remarks 116 and 119 it follows that random variables $(X_i)_{i=1}^N$ with values in $\{H, T\}$ are $\mathbb{P}$-independent iff, for every $k \in \{1, \ldots, N-1\}$,

$$\mathbb{P}(X_{k+1} = H|\sigma(X_1, \ldots, X_k)) \quad \text{is constant.}$$

Since we will normally deal with transition probabilities, this will be the most important independence criterion for us.

[117] If $\mathbb{P}(C) = 0$, then $\mathbb{P}(B|\mathcal{C})$ can be defined arbitrarily on the atom C, and it does not matter since it is a null set.

[118] Since the atoms form a finite partition which generates $\sigma(X)$.

[119] Because the atoms of $\sigma(X)$ are only $\{X = x_1\}$ and $\{X = x_2\} = \{X = x_1\}^c$.

Remark 121. We warn the reader that it is *not* true that, given 3 random variables X, Y, Z which are pairwise independent (i.e are s.t. X, Y are independent, Y, Z are independent, and X, Z are independent), then X, Y, Z are independent. For a counter-example, in the binomial model with maturity 2, let X_1, X_2 be the coin tosses and $\mathbb{P}_{\mathbf{p}}$ the probability with $\mathbf{p} = (p_1, p_2) = (\frac{1}{2}, \frac{1}{2})$, and consider the random variables

$$X := \begin{cases} 1 & \text{if } X_1 = H \\ -1 & \text{if } X_1 = T \end{cases}, \quad Y := \begin{cases} 1 & \text{if } X_2 = H \\ -1 & \text{if } X_2 = T \end{cases}, \quad Z := XY.$$

To prove the above statements, one can explicitly compute

$$\mathbb{P}(Z = 1 | X = 1) = 1/2 = \mathbb{P}(Z = 1 | X = -1),$$

which, by theorem 117 and remark 119, shows that X, Z are independent; analogously one proves that Y, Z are independent. X, Y are obviously independent (since X_1, X_2 are independent). However X, Y, Z are not independent, since $Z = XY$ is $\sigma(X, Y)$-measurable, and thus, for every Borel function f s.t. $f(Z) \in L^1$, $E(f(Z)|X, Y)$ equals $f(Z)$, which is not always a constant (for example, take $f(z) = z$ for all $z \in \mathbb{R}$), so the thesis follows from theorem 117.

Week 9

3.16 Lecture 1, Pricing and hedging fast using Markov Processes

Intuitively a process is Markov if, to estimate its future behavior, knowing the whole past is the same as knowing just its present value. Since information is represented by σ-algebras, and the (most common) estimator of a random variable Y given the σ-algebra $\mathcal{G}$ is the conditional expectation $\mathbb{E}[Y|\mathcal{G}]$, we arrive at the following definition, in which $\sigma(X_s)$ represents the information given by the present value of X, $\mathcal{F}_s$ represents the whole past, and $\mathbb{T} \subseteq [-\infty, \infty]$.

Definition 122. An adapted process $X = (X_t)_{t\in\mathbb{T}}$ on a filtered prob. space $(\Omega, \mathcal{A}, \mathcal{F}, \mathbb{P})$ is *Markov* if

$$\forall f \text{ Borel-measurable }, \forall s \leq t, \ s,t \in \mathbb{T}, \quad \mathbb{E}(f(X_t)|\mathcal{F}_s) = \mathbb{E}(f(X_t)|X_s). \tag{98}$$

It is not hard to see that there are many conditions equivalent to the one in eq. (98). For example, we could have replaced X_t for $s \leq t$ with a whole vector $(X_{t_1}, \dots, X_{t_n})$ for $s \leq t_1 \leq \dots \leq t_n$. This makes is clear that a $\mathcal{F}$-adapted process X is Markov iff the conditional law of X_t given $\mathcal{F}_s$ equals the conditional law of X_t given $\sigma(X_s)$, for all $s \leq t$; but since it is normally easier to work with the conditional expectations than with conditional probabilities, we do not often adopt this point of view. We could also have equivalently asked that

$$\forall f \text{ Borel } \forall s \leq t, \ s,t \in \mathbb{T}, \mathbb{E}(f(X_t)|\mathcal{F}_s) \text{ is } \sigma(X_s)\text{-measurable}, \tag{99}$$

i.e. that

$$\forall f \text{ Borel } \forall s \leq t, \ s,t \in \mathbb{T}, \ \exists \ g \text{ Borel s.t. } \mathbb{E}(f(X_t)|\mathcal{F}_s) = g(X_s). \tag{100}$$

Thus, if we want to prove that a process X, which only takes countably many values, is *not* Markov, we would normally show that $\mathbb{E}(f(X_t)|\mathcal{F}_s)$ is *not* constant on some set of the form $\{X_s = c\}$, and thus is not $\sigma(X_s)$-measurable (for some choice of s, t, f), thanks to lemma 79. If $\mathbb{T} = \{0, \cdots, N\}$, we could also have equivalently considered only the $s \leq t$ of the form $t = s + 1$ in eqs. (98) and (99). As eq. (99) is the easiest equivalent condition to check, it is the way one normally uses to prove that X is Markov (and, if $\mathbb{T} = \{0, \cdots, N\}$, one only needs to check eq. (99) for $t = s + 1$).

Now, consider a complete market model $(B_t, S_t)_{t\in\mathbb{T}}$, in discrete time, on a stochastic basis $(\Omega, \mathcal{A}, \mathcal{F}, \mathbb{P})$. Let $\mathbb{Q} \in \mathcal{M}(\overline{S})$ be the risk-neutral measure. If a derivative has a discounted payoff $\overline{Y}_N$ of the form $f_N(W_N)$ for some function f_N and some $\mathbb{Q}$-Markov process[120] W, by the RNFP eq. (91) we find that the AFP $(Y_n)_n$ of Y_N satisfies the following equation for $n = N - 1$

$$\overline{Y}_n = \mathbb{E}^{\mathbb{Q}}[\overline{Y}_{n+1}|\mathcal{F}_n] = \mathbb{E}^{\mathbb{Q}}[f_{n+1}(W_{n+1})|\mathcal{F}_n] = f_n(W_n) \tag{101}$$

[120] i.e. a Markov process on $(\Omega, \mathcal{A}, \mathcal{F}, \mathbb{Q})$.

for some (Borel) measurable function f_n, i.e. $\overline{Y}_n$ is of the form $f_n(W_n)$ also for $n = N-1$. We can then apply eq. (101) again with $n = N-2$ etc., so by backward induction we obtain that $\overline{Y}_n$ is of the form $f_n(W_n)$ (for some measurable function f_n) *for every* $n = 0, \dots, N$. Thus, to compute $\overline{Y}_n, f_n$ from $\overline{Y}_{n+1}$, we just need to keep track of the values of W_n, W_{n+1}, we do not need to know all the values of $(X_k)_{k\leq N}$. In particular, if given the choice between two $\mathbb{Q}$-Markov processes W, W' s.t. $\overline{Y}_N = f_N(W_N) = f'_N(W'_N)$, between W, W' one should choose the process which takes the smallest possible number of values, to minimise the time it takes to compute Y. To actually compute the prices $(\overline{Y}_n)_n$, we need to explicitly compute the pricing functions $(f_n)_n$ by backward induction, starting from f_N(which is known); we shall soon see how to do it.

Remark 123. Normally one has to take W to be a multi-dimensional process, which has the underlying S as 'one[121]' of its components, and as another component has a process C s.t. the payoff of the derivative is of the form $f_N(S_N, C_N)$. It often happens that f_N does not even depend on S_N (i.e. the payoff is actually of the form $f_N(C_N)$), but that C is not $\mathbb{Q}$-Markov while (S, C) is $\mathbb{Q}$-Markov; in this case one has to apply the above reasoning with $W = (S, C)$, and cannot apply it with $W = C$.

To prove that a process is Markov, and to compute the pricing functions $(f_n)_{n\leq N}$, normally[122] one uses the following lemma, typically taking $\mathcal{G} = \sigma(Y)$, to compute the conditional expectation.

Lemma 124 (Independence Lemma)**.** *If X a $\mathbb{R}^k$-valued, and Y is a $\mathbb{R}^n$-valued, random vector on $(\Omega, \mathcal{A}, \mathbb{P})$, $\mathcal{G} \subseteq \mathcal{A}$ is a σ-algebra, X is $\mathcal{G}$-measurable and Y is independent of $\mathcal{G}$, $f : \mathbb{R}^k \times \mathbb{R}^n \to \mathbb{R}$ is a Borel function, then*

$$\mathbb{E}(f(X,Y)|\mathcal{G}) = g(X) \quad \textit{for} \quad g(x) := \mathbb{E}(f(x,Y)), \quad g : \mathbb{R}^k \to \mathbb{R}.$$

We should stress that in the above lemma $x \in \mathbb{R}^k$ (and thus $g(x)$) is not random, whereas X, Y (and thus $g(X)$) are random. Also, although the lemma is stated for $x \in \mathbb{R}^k$, notice that to compute $g(X)$ we really only need to compute $g(x)$ for all x in the image[123] of X. In particular, if X only takes[124] (finitely or) countably many values $(x_k)_k$, we only need to compute $(g(x_k))_k$. Let us now illustrate how to use the above lemma.

Example 125. Given X, Y independent and uniformly distributed in $[0, 1]$, we can com-

[121] Here S could be multi-dimensional.

[122] If the (joint) law of (X, Y) is known, there is another method commonly used to compute $\mathbb{E}(X|Y)$, but we will not need it in this class.

[123] To be precise, since X, being a random variable, is only defined up to $\mathbb{P}$-null sets, we should look at the $\mathbb{P}$-essential image of X, defined as the support of the law of X, i.e. the intersection of all closed sets $C \subseteq \mathbb{R}^k$ s.t. $X \in C$ $\mathbb{P}$ a.s..

[124] Meaning $\mathbb{P}(X = x_k) > 0, \mathbb{P}(X \notin \cup_k \{x_k\}) = 0$.

pute $\mathbb{E}(X \wedge Y|X)$: it equals $g(X)$, where

$$g(x) = \mathbb{E}(x \wedge Y) = \mathbb{E}(Y 1_{\{Y \le x\}} + x 1_{\{Y > x\}}) = \int_0^1 y 1_{\{y \le x\}} dy + x \int_0^1 1_{\{y > x\}} dy$$
$$= \int_0^x y dy + x \int_x^1 1 dy = \frac{x^2}{2} + x(1-x) = x - \frac{1}{2}x^2, \quad \forall x \in \mathbb{R}.$$

Thus $\mathbb{E}(X \wedge Y|X) = g(X) = X - \frac{1}{2}X^2$.

Corollary 126. *If a $\mathcal{F}$-adapted process W satisfies $W_{k+1} = f_k(W_k, X_{k+1})$ for each $k \in \mathbb{Z}$, where f_k is some Borel function and X_{k+1} is independent of $\mathcal{F}_k$, then W is Markov and $\mathbb{E}(f(W_{k+1})|\mathcal{F}_k) = g(W_k)$ for all $k \in \mathbb{Z}$, where*

$$g(w) := \mathbb{E}[f(f_k(w, X_{k+1}))] \tag{102}$$

Proof. For every k we have that

$$\mathbb{E}(f(W_{k+1})|\mathcal{F}_k) = \mathbb{E}((f \circ f_k)(W_k, X_{k+1})|\mathcal{F}_k) = g(W_k)$$

where g is the Borel function in eq. (102), so W is Markov. □

Obviously corollary 126 also applied if the time index $\mathbb{Z}$ is replaced by any subset thereof. Let use illustrate the use of corollary 126 in proving that a process is Markov.

Example 127. Assume that in the binomial model $S_{n+1} = f_n(S_n, X_{n+1})$, where X_{n+1} is the $(n+1)^{th}$-coin toss and f_n is a (Borel) function (this happens for example when the up and down factors u_n, d_n are deterministic). Since S is $\mathcal{F}$-adapted and X_{n+1} is independent of $\mathcal{F}_n$ under $\mathbb{P}$ for all n, S is a Markov process under $\mathbb{P}$. Analogously, S is Markov under the risk neutral measure $\mathbb{Q}$ if X_{n+1} is independent of $\mathcal{F}_n$ under $\mathbb{Q}$ for all n (i.e. if the $(X_n)_n$ are $\mathbb{Q}$-independent); this happens if and only if $\tilde{P}_n$ is deterministic[125] for all n, because of remark 120.

Remark 128 (Hedging fast using Markov Processes). Assume we are in the binomial model and choose an adapted W process s.t. $\overline{Y}_N = f_N(W_N)$, $W_{k+1} = h_k(W_k, X_{k+1})$ and $S_k = s_k(W_k)$ for each k, where h_k, s_k are some Borel functions and $(X_k)_k$ are the coin tosses. If[126] W is $\mathbb{Q}$-Markov then $\overline{Y}_n = f_n(W_n)$ for all $n \le N$ for some $(f_n)_n$. Since the delta-hedging formula gives

$$G_n(\omega) = G_n(\omega(n)) = \frac{\overline{Y}_{n+1}(\omega(n)H) - \overline{Y}_{n+1}(\omega(n)T)}{\overline{S}_{n+1}(\omega(n)H) - \overline{S}_{n+1}(\omega(n)T)}$$

we get

$$G_n(\omega(n)) := \frac{f_{n+1}(W_{n+1})(\omega(n)H) - f_{n+1}(W_{n+1})(\omega(n)T)}{s_{n+1}(W_{n+1})(\omega(n)H) - s_{n+1}(W_{n+1})(\omega(n)T)}$$

[125] Though it is allowed to depend on n, it cannot be random, i.e., it cannot depend on ω.
[126] Which, by corollary 126 happens in particular if the $(X_n)_n$ are $\mathbb{Q}$-independent, i.e. if $\tilde{P}_n$ is deterministic.

and since

$$W_{n+1}(\omega(n)H) = h_n(W_n(\omega(n)), H), \quad W_{n+1}(\omega(n)T) = h_n(W_n(\omega(n)), T)$$

we find that $G_n(\omega(n)) = g_n(W_n(\omega(n))$ for some Borel function g_n, which can be calculated explicitly as

$$g_n(w) := \frac{f_{n+1}(h_n(w,H)) - f_{n+1}(h_n(w,T))}{s_{n+1}(h_n(w,H)) - s_{n+1}(h_n(w,T))}. \tag{103}$$

Thus, not only the price $\overline{Y}_n$ but also the hedging strategy $G_n = g_n(W_n)$ can be quickly computed by looking only at the values of W_n, and by calculating the portfolio function g_n, which only needs to be calculated for w in the 'image'[127] of W_n.

3.17 Lecture 2, Example of Markov pricing

Let us illustrate with an example the method described in the previous section by computing the price V a (floating) lookback put option, i.e. the derivative which has payoff $V_N := M_N - S_N$ at maturity N, where $M_n := \max_{i=0,\dots,n} S_i$ is the running maximum of the price S of the underlying. We describe $S = (S_n)_{n=0}^N$ with a N-period binomial model with constant parameters $S_0, u, d, r > 0$, which are assumed to satisfy the no-arbitrage condition $0 < d < 1 + r < u$.

Notice that the (discounted) payoff of the derivative in question can be written as

$$\overline{V}_N = f_N(S_N, M_N), \text{ for } \quad f_N(s,m) := \frac{1}{(1+r)^N}(m-s).$$

This suggests to consider the process $W := (S, M)$, and to try to prove that it is $\mathbb{Q}$-Markov. As we saw in eq. (102), an easy way to do that would be to prove the existence of h_n s.t. $W_{n+1} = h_n(W_n, X_{n+1})$, by computing it explicitly. To this end, notice that

$$\frac{S_{n+1}}{S_n} = q(X_{n+1}), \text{ where } \quad q(x) := \begin{cases} u, & \text{if } x = H \\ d, & \text{if } x = T \end{cases},$$

and this allows to write

$$S_{n+1} = S_n q(X_{n+1}), \quad M_{n+1} = M_n \vee S_{n+1} = M_n \vee (S_n q(X_{n+1}))$$

from which we conclude that indeed $W_{n+1} = h_n(W_n, X_{n+1})$ if we take

$$h_n(s, m, x) := (sq(x), m \vee (sq(x))).$$

We can then apply the Markov pricing method to the problem at hand, as follows. Since

$$\overline{V}_n = \mathbb{E}^{\mathbb{Q}}[\overline{V}_{n+1}|\mathcal{F}_n] = \mathbb{E}^{\mathbb{Q}}[f_{n+1}(W_{n+1})|\mathcal{F}_n] = f_n(W_n),$$

[127] See the discussion right after lemma 124.

and $\overline{V}_N = f_N(S_N, M_N)$ for some function f_N, we conclude by backward induction that $\overline{V}_n = f_n(W_n)$ *for every* n, and to compute the price V of the derivative the pricing functions f_n can (and should) be explicitly calculated by backward induction as follows. By the independence lemma $\mathbb{E}^{\mathbb{Q}}[f(W_{n+1})|\mathcal{F}_n] = g(W_n)$, where

$$g(s,m) := \mathbb{E}^{\mathbb{Q}}[f(h_n(s,m,X_{n+1}))] = \tilde{p}f(su, m \vee (su)) + \tilde{q}f(sd, m \vee (sd)),$$

where $\tilde{p} := \mathbb{Q}(X_{n+1} = H)$, $\tilde{q} := 1 - \tilde{p}$; notice that

$$\mathbb{Q}(X_{n+1} = H) = \mathbb{E}^{\mathbb{Q}}[\mathbb{Q}(X_{n+1} = H|\mathcal{F}_n)],$$

and so in the present setting $\tilde{p}$ equals

$$\tilde{p}_n := \mathbb{Q}(X_{n+1} = H|\mathcal{F}_n) = \frac{(1+r)-d}{u-d},$$

since the latter is deterministic. In summary, the pricing functions are given by

$$\begin{aligned} f_N(s,m) :=& \frac{1}{(1+r)^N}(m-s), \\ f_n(s,m) :=& \tilde{p}f_{n+1}(su, m \vee (su)) + \tilde{q}f_{n+1}(sd, m \vee (sd)), \qquad n = 0, \ldots, N-1. \end{aligned}$$

Finally, the replicating strategy $G_n = g_n(S_n, M_n)$ is given by eq. (103), where s_n is the function $s_n(w) = s_n(s,m) := s$ (since it satisfies $S_n = s_n(W_n)$), and so

$$g_n(s,m) := \frac{f_{n+1}(su, m \vee (su)) - f_{n+1}(sd, m \vee (sd))}{su - sd}, \quad n = 0, \ldots, N-1.$$

4 Continuous time models

EVERYTHING FROM NOW ON IS NOT EXAMINABLE

We will conclude this module with a brief introduction to the most important model in continuous time (the Black and Scholes model), and show how it is obtained by taking limits of the multi-period binomial model. We will also show how the risk-neutral pricing method can be applied in this setting; we will *not* discuss replication in the BS (Black and Scholes) model, since this would require very sophisticated mathematics (namely, stochastic calculus).

4.1 Lecture 3, From the Binomial to the Black and Scholes model

Suppose now we want to consider models with price movements at arbitrary dates and of arbitrary size. This is not really because it makes the models more realistic: after all, taking a discrete-time model with extremely small time intervals based on a probability space made of really many points, one can approximate arbitrarily well any continuous-time model. The real reason for wanting to work in continuous time here is the same as for in classical physics: we want to use the power of differential calculus and its rich bag of tools (integrals, ODEs, the fundamental theorem of calculus...) to be able to make explicit calculations which are just not possible when working in discrete-time (using sums, difference equations, ...). The only difference is that, when working with stochastic processes in continuous time, one deals with a much more complicated type of calculus (called *stochastic calculus*), and its tools (stochastic integrals, SDEs (Stochastic Differential Equations), Ito's formula, ...).

So let use as time index the interval $[0, T]$, and consider a price process $(B_t)_{t\in[0,T]}$ for the bond and $(S_t)_{t\in[0,T]}$ for the underlying. How should we model (B, S)? It is quite intuitive that B, S should be continuous[128] processes; whether this can actually be concluded applying statistical considerations to market data is still the subject of debate. To determine what model we could plausibly choose, let us see what happens when taking the limit of binomial models with constant coefficients in the most intuitive way possible. For $N = 1, 2, \dots$ let $\Delta := T/N$, and consider as discrete-time index for a binomial model $(B_t^N, S_t^N)_{t\in\pi^N}$ the following partition π^N, which is a discretized version of $[0, T]$

$$\pi^N := \{0, \Delta, 2\Delta, \cdots, N\Delta\} = \{t_i^N\}_{i=0}^N, \qquad \text{where } t_i^N := i\Delta.$$

In order to have $\pi^N \subseteq \pi^{N+1}$, one should only consider N of the form $N = 2^i$ for some $i \in \mathbb{N}$, though we won't indicate that in our notation, to keep it lighter. We should also extend our model $(B_t^N, S_t^N)_{t\in\pi^N}$ to all $t \in [0, T]$; we can do so by declaring $t \mapsto B_t^N, S_t^N$ to be affine in each interval $[t_i^N, t_{i+1}^N]$, so that $(B_t^N, S_t^N)_{t\in[0,T]}$ is a continuous process. For each N, for simplicity we take the parameters r, u, d of the model (B_t^N, S_t^N) to be as simple as possible: they should be constants in t, ω, but they should depend on N to have a non-trivial limit (B, S) of (B^N, S^N). How exactly should the parameters be chosen to depend on N it is not clear for now. The right dependence for r^N is easy to guess: if in

[128] $(X_t)_t$ is continuous if $t \mapsto X_t(\omega)$ is continuous $\mathbb{P}$ a.e. ω.

discrete time the bank account has an (un-compounded) interest rate r per unit time, i.e. investing 1 at time $i\Delta$ we get back $(1 + r\Delta)$ at time $(i + 1)\Delta$, then the interest rate in the model (B^N, S^N) should be given by r times the step of the partition π_N, i.e. $r^N = rT/N$. This dependence in N gives the following reasonable result: compounding the interest we get $B^N_T = (1 + rT/N)^N$, and so $B^N_T \to \exp(rT) = B_T$ for $N \to \infty$. Thus taking $r^N := rT/N$ results in the limit B of B^N to be given by the model of a bank account B with constant short rate r.

How should be choose d^N, u^N to get S^N to convergence in law (a.k.a. in distribution) to a non-trivial process S, and more importantly, what will then be S? Notice that the $R^N_i := S^N_{t^N_i}/S^N_{t^N_{i-1}}$ are IID under the physical measure $\mathbb{P}$, and $S^N_{t^N_k} = S_0 \Pi^k_{i=1} R^N_i$, and it is natural to want to transform the product of IID into a sum, so that we can apply the CLT (Central Limit Theorem). So, we consider the process $X^N := \ln(S^N)$, so that

$$X^N_{t^N_k} = X^N_0 + \textstyle\sum_{i=1}^k Y^N_i \quad \text{with } Y^N_i := \ln(R^N_i), \quad X^N_0 := \ln(S_0).$$

To have X^N converge, we plan to apply the following slight extensions of the CLT.

Lemma 129. *Let $\left(Z^n_j\right)_j$ be IID, so $Z^n_j \sim Z^n$ for all j. Assume $Z^n \in L^2(\mathbb{P})$, $E(Z^n) \to m \in \mathbb{R}$ and $\mathrm{Var}\,(Z^n) \to \sigma^2$ as $n \to \infty$, where $0 < \sigma^2 < \infty$. Then $S_n := \frac{\sum_{j=1}^n (Z^n_j - m)}{\sqrt{n}}$ converges in distribution to the normal random variable $S \sim \mathcal{N}(0, \sigma^2)$*

The proof of the previous lemma is identical to the most commonly used proof of the CLT (the one with characteristic functions); the statement is however more general, in that the Z^n_j are allowed to also depend on n, and instead of asking that $Z_j \sim Z$ satisfy

$$E(Z) = m, \mathrm{Var}\,(Z) = \sigma^2$$

we ask that $Z^n_j \sim Z^n$ satisfy

$$E(Z^n) \to m, \mathrm{Var}\,(Z^n) \to \sigma^2 \text{ as } n \to \infty.$$

Consider now the increment $X^N_{t^N_k} - X^N_{t^N_{k-1}} = Y^N_k$ of X during the interval $[t^N_{k-1}, t^N_k]$, which has length $t^N_k - t^N_{k-1} = \Delta = T/N$. Assume that u^N, d^N are chosen[129] so that the limits of the expected value and variance of each increment of X per unit time exist and are non-trivial, i.e. for all i

$$\exists \lim_N \frac{N}{T} \mathbb{E}^{\mathbb{P}}(Y^N_i) = \mu \in \mathbb{R}, \quad \exists \lim_N \frac{N}{T} var^{\mathbb{P}}(Y^N_i) = \sigma^2 \in (0, \infty),$$

and that the convergence of $\frac{N}{T} Y^N_i$ to μ is faster than $1/\sqrt{N}$, i.e.

$$\frac{1}{\sqrt{N}} \mathbb{E}^{\mathbb{P}}(\frac{N}{T} Y^N_i - \mu) \to 0.$$

[129] We don't show how to choose such u^N, d^N for reasons of brevity. In the literature one can find several different possible expressions of u^N, d^N.

In terms of the rv $Z_i^N := \frac{1}{\sqrt{N}}(NY_i^N - T\mu)$ we can write our assumptions as

$$\mathbb{E}^{\mathbb{P}}(Z_i^N) \to 0, \quad var^{\mathbb{P}}(Z_i^N) = var^{\mathbb{P}}(\sqrt{N}Y_i^N) \to \sigma^2 T \quad \text{as } N \to \infty. \tag{104}$$

Then, writing

$$X_T^N - X_0^N - T\mu = \sum_{i=1}^N (Y_i^N - T\mu/N) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{1}{\sqrt{N}}(NY_i^N - T\mu) = \frac{1}{\sqrt{N}} \sum_{i=1}^N Z_i^N$$

we can apply lemma 129 and get that

$$X_T^N - X_0^N - T\mu = \frac{1}{\sqrt{N}} \sum_{i=1}^N Z_i^N \overset{\text{LAW}}{\longrightarrow} \sigma W_T, \quad W_T \sim \mathcal{N}(0, T). \tag{105}$$

So $X_T^N \to X_0 + \mu T + \sigma W_T =: X_T$, which implies $S_T^N \to S_0 \exp(\mu T + \sigma W_T) =: S_T$ for a $W_T \sim \mathcal{N}(0, T)$. This only determined the law of S_T; analogously we get that S_t^N converges in law to $S_0 \exp(\mu t + \sigma W_t) =: S_t$ for a $W_t \sim \mathcal{N}(0, t)$; this is not enough, as we need to determine the law of the whole process S, i.e. explain how S moves between different times. To do so, we reason analogously, and if $s, t \in \pi^N \subseteq \pi^{N+1}, s \leq t$, for $N \to \infty$ we get

$$X_t^N - X_s^N - \mu(t-s) \longrightarrow \sigma \cdot (W_t - W_s), \quad W_t - W_s \sim \mathcal{N}(0, t-s).$$

If $r, s, t \in \pi^N, r \leq s \leq t$ then $X_s^N - X_r^N$ and $X_t^N - X_s^N$ are independent[130], i.e.

$$\mathbb{E}[f(X_s^N - X_r^N) g(X_t^N - X_s^N)] = \mathbb{E}[f(X_s^N - X_r^N)]\, \mathbb{E}[g(X_t^N - X_s^N)], \quad \text{for all Borel } f, g,$$

and taking $N \to \infty$ shows that $X_t - X_s$ and $X_s - X_r$ are independent, so $W_t - W_s$ and $W_s - W_r$ are independent. Working analogously with arbitrary times $0 \leq s_0 < s_1 < \cdots < s_k \leq T, k \in \mathbb{N}$, we could show that W satisfies

$$(W_{s_1} - W_{s_0}, W_{s_2} - W_{s_1}, \cdots, W_{s_k} - W_{s_{k-1}}) \text{ are independent.} \tag{106}$$

Such a process W is said to have *independent increments.*

In summary, we are thus lead to considering the following definition for the process W and the limiting model (B, S).

Definition 130. A (standard, one dimensional) BM (*Brownian Motion*) is a process $W = (W_t)_{t \geq 0}$ and s.t.:

0. $W_0 = 0$.

1. W has independent increments.

2. $W_t - W_s \sim \mathcal{N}(0, (t-s))$ for $0 \leq s < t$.

[130] Since $X_s^N - X_r^N = \sum_{i \in I} Y_i^N$ and $X_t^N - X_s^N = \sum_{i \in J} Y_i^N$ for some disjoint I, J, and the $(Y_i^N)_i$ are independent.

3. W is continuous (i.e. $t \mapsto W_t(\omega)$ continuous $\mathbb{P}$ a.e. ω).

Remark 131. A process $W = (W^i)_{i=1}^n$ with values in $\mathbb{R}^n$ is a called a *n-dimensional BM* if its components $W^1, \dots, W^n$ are independent processes[131], and each component is a 1-dimensional BM.

Remark 132 ($\mathcal{F}$-Brownian Motion). Normally, one takes $\mathcal{F}$ to be the natural filtration of BM, i.e. $\mathcal{F}_t = \mathcal{F}_t^W := \sigma(\{W_s : s \in [0,t]\})$ for all t. It is sometimes useful to consider a more general filtration $\mathcal{F}$. Then we say that W is a $\mathcal{F}$-BM if it is a BM and

$$W \text{ is } \mathcal{F}\text{-adapted}, \quad W_t - W_s \text{ is independent of } \mathcal{F}_s \text{ for every } t \geq s \geq 0. \tag{107}$$

Trivially eq. (107) implies item 1 in definition 130, and is equivalent to it if $\mathcal{F} = \mathcal{F}^W$ (so W is a BM iff it is a $\mathcal{F}^W$-BM).

Remark 133. If item 0 in definition 130 is replaced by

$$W_0 = X$$

where X is any random variable independent from W, then the process $B_t := X + W_t, t \geq 0$, which satisfies all items in definition 130 but for item 0, is called a BM started at X.

Definition 134. The BS (*Black-Scholes*) model considers a bank account B and one stock S modelled as

$$B_t := \exp(rt), \quad S_t := S_0 \exp(\mu t + \sigma W_t) \quad \text{for } S_0, \sigma > 0, r > -1, \mu \in \mathbb{R}. \tag{108}$$

where W is a (one-dimensional) Brownian Motion, and the natural filtration $\mathcal{F}^W$ of BM. Such process S is called a GBM (*Geometric Brownian Motion*).

Remark 135. The process $L := \log(S)$ is Gaussian and has independent increments, since if $s < t < u$ then

$$L_t - L_s = \log(S_t) - \log(S_s) = \log(S_t/S_s) = \mu(t-s) + \sigma(W_t - W_s).$$

In particular, $\dfrac{S_t}{S_s}$ and $\dfrac{S_u}{S_t}$ are independent; notice that however $S_t - S_s$ and $S_u - S_t$ are **not** independent!

[131] By definition, processes $\{X^i\}_{i \in I}$ are independent if they generate independent σ-algebras $\sigma(X^i) := \sigma(\{X_t^i\}_{t \geq 0}), i = 1, \dots, n$.

Week 10

4.2 Lecture 1, The EMM in the BS model

The previous calculations of convergence of the binomial model (B^N, S^N) to the BS model (B, S) were done with convergence in law under the underlying probability $\mathbb{P}$, which describes how likely is each market scenario; such $\mathbb{P}$ is called the *physical measure*. Instead, we now look at the convergence in law of (B^N, S^N) under its EMM $\mathbb{Q}^N$ to determine the law of (B, S) under its EMM $\mathbb{Q} =: \widetilde{\mathbb{P}}$. Each R_i^N takes[132] values u^N, d^N under $\mathbb{Q}^N$, and the $(R_i^N)_i$ are IIDs also under $\mathbb{Q}^N$, since the risk-neutral transition probabilities

$$\tilde{p}^N = \frac{(1 + r^N) - d^N}{u^N - d^N}$$

are constants. Thus, we will still be able to carry out the same calculations as before, and get that the limiting model is $S_t := S_0 \exp(\tilde{\mu}t + \tilde{\sigma}\tilde{W}_t)$, where $\tilde{W}$ is a Brownian Motion under $\mathbb{Q}$; what changes now is that, since the expectation and variance of Z_i^N under $\mathbb{P}^N$ is not necessarily the same as under $\mathbb{Q}^N$, the values of $\tilde{\mu}, \tilde{\sigma}$ may be different from μ, σ. It turns out that $\tilde{\sigma} = \sigma$; this important fact can be proved[133] by showing that, if for some $\mathbb{P}' \sim \mathbb{P}, a > 0, b \in \mathbb{R}$ the process $Y_t := a(W_t + bt)$ is a $\mathbb{P}'$-BM, then $a = 1$ (because the quadratic variation $[Y]$ of Y is $a^2 t$). As for $\tilde{\mu}$, it is determined by using the fact that $\overline{S} = S/B$ is a $\mathbb{Q}$-martingale, i.e. $\overline{S}_s$ equals $\mathbb{E}^{\mathbb{Q}}[\overline{S}_t|\mathcal{F}_s]$ for all $s \leq t$. Since $S_t := S_0 \exp(\tilde{\mu}t + \sigma\tilde{W}_t)$, we have that

$$\overline{S}_t = S_0 \exp((\tilde{\mu} - r)t + \sigma\tilde{W}_t) = A_t \mathcal{E}_t(\sigma\tilde{W}), \text{ for } A_t := S_0 \exp\left(\left(\tilde{\mu} - r + \frac{\sigma^2}{2}\right)t\right). \quad (109)$$

Since $\mathcal{E}(\sigma\tilde{W})$ is a martingale and A_t is deterministic we have

$$\mathbb{E}^{\mathbb{Q}}[\overline{S}_t|\mathcal{F}_s] = \mathbb{E}^{\mathbb{Q}}[A_t\mathcal{E}_t(\sigma\tilde{W})|\mathcal{F}_s] = A_t\mathbb{E}^{\mathbb{Q}}[\mathcal{E}_t(\sigma\tilde{W})|\mathcal{F}_s] = A_t\mathcal{E}_s(\sigma\tilde{W}),$$

whereas applying eq. (109) with $t = s$ we find $\overline{S}_s = A_s\mathcal{E}_s(\sigma\tilde{W})$. Thus, $\overline{S}$ is a $\mathbb{Q}$-martingale iff $A_t = A_s$ for all $s \leq t$, i.e iff $\tilde{\mu} = r - \sigma^2/2$. In summary, under appropriate choices of u^N, d^N we will get that S^N will converge in law under $\mathbb{Q}^N$, and its limit in law S will be given by

$$S_t = S_0 e^{rt}\mathcal{E}_t(\sigma\widetilde{W}) = \exp\left((r - \frac{\sigma^2}{2})t + \sigma\widetilde{W}_t\right), \quad (110)$$

where $\widetilde{W}$ is a BM under the EMM $\mathbb{Q}$. Thus, eq. (110) provides us with the law of S under $\mathbb{Q}$, which is what we need to apply the RNPF.

---

[132] Meaning $\mathbb{Q}^N(R_i^N = x) > 0$ if $x \in \{u^N, d^N\}$, and $\mathbb{Q}^N(R_i^N \notin \{u^N, d^N\}) = 0$.

[133] It could be proved also by explicitly choosing appropriate u^N, d^N and doing the calculations, but this is tedious and not as insightful.

4.3 Lecture 2, Pricing the call in the binomial model

Since the binomial model $(B_t, S_t)_{t=0,1,\ldots,N}$ has a unique EMM $\mathbb{Q}$, the price at time $n \le N$ of the call with strike K and maturity N satisfies $\overline{C}_n = \mathbb{E}^{\mathbb{Q}}[\overline{C}_N|\mathcal{F}_n]$, and so

$$C_n = \mathbb{E}^{\mathbb{Q}}[\frac{B_n}{B_N}(S_N - K)^+|\mathcal{F}_n] = \frac{1}{(1+r)^{N-n}}\mathbb{E}^{\mathbb{Q}}[\left(S_n\frac{S_N}{S_n} - K\right)^+|\mathcal{F}_n].$$

Since $S_N/S_n = \Pi_{i=n+1}^N R_i$ is independent of $\mathcal{F}_n$ and S_n is $\mathcal{F}_n$-measurable, by applying the independence lemma we get that $C_n = c_n(S_n) := c(n, S_n)$ where

$$c(n,x) := \frac{1}{(1+r)^{N-n}}\mathbb{E}^{\mathbb{Q}}[(x\Pi_{n+1}^N R_i - K)^+]. \tag{111}$$

Since the $(R_i)_i$ are IID under $\mathbb{Q}$ and $B_n = (1+r)^n$, c can be computed as

$$c(n,x) = \frac{1}{(1+r)^{N-n}}\sum_{j=0}^{N-n}\binom{N-n}{j}\tilde{p}^j(1-\tilde{p})^{N-n-j}(xu^jd^{N-n-j} - K)^+. \tag{112}$$

4.4 Lecture 3: Pricing the call in the BS model

Since the BS model has a unique EMM $\mathbb{Q}$, the price at time $t \le T$ of the call with strike K and maturity T in the BS model satisfies $\overline{C}_t = \mathbb{E}^{\mathbb{Q}}[\overline{C}_T|\mathcal{F}_t]$, and so

$$C_t = \mathbb{E}^{\mathbb{Q}}[\frac{B_t}{B_T}(S_T - K)^+|\mathcal{F}_t] = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}[\left(S_t\frac{S_T}{S_t} - K\right)^+|\mathcal{F}_t], \tag{113}$$

Since the binomial models (B^N, S^N) converge in law under $\mathbb{Q}^N$ to the BS model (B, S) under $\mathbb{Q}$, the price C_t^N of the call in the model (B^N, S^N) converges to the price C_t in the BS model. It would then be possible, but messy, to compute C_t by taking the limit of C_t^N. Let us instead compute C_t directly from eq. (113), emulating our discrete-time calculations. From eq. (110) we get that

$$S_T/S_t = \exp((r - \sigma^2/2)(T-t) + \sigma(\widetilde{W}_T - \widetilde{W}_t)),$$

which shows that S_T/S_t is independent of $\mathcal{F}_t$ (under $\mathbb{Q}$). Since S_t is $\mathcal{F}_t$-measurable, the independence lemma gives that $C_t = c(t, S_t)$, where

$$c(t,x) := e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\left[\left(x\frac{S_T}{S_t} - K\right)^+\right]. \tag{114}$$

To compute $c(t,x)$, define

$$Y := -\frac{\widetilde{W}_T - \widetilde{W}_t}{\sqrt{T-t}}, \quad h(y) := h(t,x,y) := x\exp\left(-\sigma\sqrt{T-t}y + (r - \frac{\sigma^2}{2})(T-t)\right)$$

so that

$$Y \underset{\text{under } \mathbb{Q}}{\sim} \mathcal{N}(0,1), \quad x\frac{S_T}{S_t} = h(t,x,Y) \quad c(t,x) = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\left[(h(t,x,Y)-K)^+\right]. \tag{115}$$

Since c and h only depend on (t,T) as a function of $T-t$, we will express this dependence using the variable $\tau := T-t$. To compute c, we need to identify when the payoff of the call option is non-zero in terms of Y, and so we write

$$X := x\frac{S_T}{S_t} - K \geq 0 \iff -\sigma\sqrt{\tau}Y + (r - \frac{\sigma^2}{2})\tau = \ln\left(\frac{S_T}{S_t}\right) \geq \ln\left(\frac{K}{x}\right) = -\ln\left(\frac{x}{K}\right),$$

which holds iff

$$Y \leq d_- := d_-(\tau,x) := \frac{1}{\sigma\sqrt{\tau}}\left(\ln\left(\frac{x}{K}\right) + \left(r - \frac{\sigma^2}{2}\right)\tau\right),$$

and thus $\{X \geq 0\} = \{Y \leq d_-(\tau,x)\}$. Using that $X^+ = X1_{\{X\geq 0\}}$ and $X = h(Y) - K$, it follows that

$$\widetilde{\mathbb{E}}((x\frac{S_T}{S_t} - K)^+) = \widetilde{\mathbb{E}}(1_{\{Y\leq d_-\}}(h(Y)-K)).$$

Since $Y \sim \mathcal{N}(0,1)$ has density $\phi(y) := \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}$ we get that

$$c(t,x) = e^{-r\tau}\int_{-\infty}^{d_-}\left(x\exp\left(-\sigma\sqrt{\tau}y + (r - \frac{\sigma^2}{2})\tau\right) - K\right)\phi(y)dy.$$

In terms of the CDF of the standard Gaussian

$$\mathcal{N}(x) := \int_{-\infty}^{x}\frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}dy, \tag{116}$$

which cannot be expressed by an analytic formula but can be calculated numerically very fast, we can now write $c(t,x) = A - B$, with

$$A := \int_{-\infty}^{d_-} x\exp\left(-\frac{1}{2}(y+\sigma\sqrt{\tau})^2\right)\frac{dy}{\sqrt{2\pi}}, \qquad B := e^{-r\tau}K\mathcal{N}(d_-).$$

To compute A, change variable $z := y + \sigma\sqrt{\tau}$, so $dz = dy$ and

$$A = \int_{-\infty}^{d_+} x\frac{e^{-z^2/2}}{\sqrt{2\pi}}dz = x\mathcal{N}(d_+), \text{ where } d_+(\tau,x) := d_-(\tau,x) + \sigma\sqrt{\tau}.$$

Notice that both d_+ and d_- can be conveniently expressed by the formula

$$d_\pm := d_\pm(T-t,x) := \frac{1}{\sigma\sqrt{T-t}}\left(\log\left(\frac{x}{K}\right) + \left(r \pm \frac{\sigma^2}{2}\right)(T-t)\right). \tag{117}$$

In summary we obtained the celebrated Black and Scholes option pricing formula

$$c(t,x) = x\mathcal{N}(d_+) - e^{-r(T-t)}K\mathcal{N}(d_-), \tag{118}$$

for the price of call option at time t in the Black-Scholes model if $S_t = x$; here $\mathcal{N}$ and $d_\pm$ are defined in eqs. (116) and (117). Notice that $c, d_\pm$ only depend on (t,T) via the time to maturity $\tau = T-t$.

4.5 Lecture 4, Hedging the call in the binomial and BS models

In the binomial model, the replicating strategy $G = (G_n)_n$ for the call option is given by the delta-hedging formula as

$$G_{n-1} = \frac{c(n, S_{n-1}u) - c(n, S_{n-1}d)}{S_{n-1}u - S_{n-1}d}, \tag{119}$$

Since delta-hedging formula eq. (119) expresses the trading strategy as the slope of the pricing function $c_n(\cdot)$ (computed in eq. (112)) between the two possible values of S_n, one could guess that in continuous time the hedging strategy should be

$$G_t = \left(\frac{\partial c}{\partial x}\right)(t, S_t); \tag{120}$$

this turns out to be true, but its proof requires technical tools which are beyond the scope of this class. Since for $b, s \geq 0, K \in \mathbb{R}$ the function $x \mapsto b(sx - K)^+$ is positive, increasing and convex, from eqs. (111) and (114) it follows that $c(n, \cdot), c(t, \cdot)$ are positive, increasing and convex. Since $c(n, \cdot), c(t, \cdot)$ are increasing, eqs. (119) and (120) show that $G_n, G_t \geq 0$, i.e. the replicating strategy of the call option in the binomial and Black and Scholes models does not involve short-selling.

4.6 Lecture 5, The Greeks

The partial derivatives of the option pricing function c given by eq. (118) with respect to its multiple arguments are called the *Greeks*, because they are traditionally denoted with the following Greek[134] letters. The derivatives of c are important because they describe how c changes when its arguments change.

1. The dependence on the price of S is measured by the *Delta* $\Delta := \partial c/\partial x$.
2. Second-order effects on x involve the *Gamma* $\Gamma := \partial^2 c/\partial x^2$.
3. The time-dependence is given by *Theta* $\Theta := \partial c/\partial t$.
4. Volatility dependence is given by *Vega* $\nu := \partial c/\partial \sigma$.
5. The sensitivity to interest rates is given by *Rho* $\rho := \partial c/\partial r$.

Using the explicit formula eq. (118) we could painstakingly compute the Greeks explicitly; as this is purely an exercise in calculus, we don't do it here, and we just mentioned that the most important Greek (the Delta) is given by the nice formula

$$\Delta = \mathcal{N}(d_+) > 0.$$

To easily remember this formula, you should remember that while d_+, d_- depend on $(T - t, x)$, and thus $\Delta = \partial c/\partial x$ must be computed from eq. (118) using the chain rule,

[134] Other than Vega, which comes from Spanish, though normally people use the Greek letter ν as its symbol.

it turns out that their effects cancel out, so $\partial c/\partial x$ equals the value that it would take if d_+, d_- did not depend on x.

Using the explicit formulas for the Greeks one can check that, for the call option,

$$Delta > 0, \quad Gamma > 0, \quad Theta < 0, \quad Vega > 0.$$

The fact that $Delta > 0$ is intuitive: the larger $S_t = x$, the larger the payoff of the call $(xS_T/S_t - K)^+$, so $x \mapsto c(t,x)$ is increasing (this follows from the domination principle, and also from the RNFP). As we already mentioned, this implies in particular that replicating a call involves no short-selling. Analogously, since $c(\cdot, T) = (\cdot - K)^+$ is convex, the domination principle shows that $c(t, \cdot)$ is convex, and so $\Gamma \geq 0$ for the call option.

Also $Vega > 0$ is intuitive: a more volatile stock means that the outcome is more uncertain, and so the insurance against adversity provided by the call option becomes more valuable (you can say that *options like volatility*).

Suppose at time t we want to take a long position in the call option and hedge it. We then buy one call option, spending $c(t, S_t)$, and hedge it by shorting $\Delta_t = \partial_x c(t, S_t)$ shares at time t, which generates an income of $\Delta_t S_t$, so that our wealth in the bank is

$$M := M(t, S_t) := S_t \cdot \partial_x c(t, S_t) - c(t, S_t).$$

Notice that substituting the formulas for c and $\partial_x c$ we get that $M = e^{-r(T-t)} K \mathcal{N}(d_-)$, and so $M > 0$, so to hedge a long position in a call option one never needs to borrow any money. If $x := S_t$, and writing $c(x), M(x)$ for $c(t,x), M(t,x)$, our total wealth is

$$V(x) := V(t,x) = c(x) - x\partial_x c(x) + M(x) = 0.$$

If S_t were to suddenly jump to y, then $V(x)$ would become

$$V(y) := c(y) - y\partial_x c(x) + M(x) = c(y) - (c(x) + (y-x)\partial_x c(x)), \tag{121}$$

since the amount of money in the bank remains $M(x)$, and the amount of shares you own remains $\partial_x c(x)$, but the call option and the underlying change value from $c(x), x$ to $c(y), y$. Geometrically, the quantity in eq. (121) represents the difference between the value at the point y of a strictly convex function c, and the value at y of the line tangent to c at x; in particular V is positive and $V(y) = 0$ iff $y = x$. Thus V achieves its unique minimum at $y = x$, and in particular $(\partial_y V)(x) = 0$; moreover $\partial^2_{y^2} V = \partial^2_{y^2} c \geq 0$. Our portfolio is then said to be *delta-neutral*, *long gamma*, because

$$(\partial_y V)(x) = 0, \quad \partial^2_{y^2} V > 0.$$

Since $V(y) > 0$ for any $y \neq x$, we benefit from any instantaneous jumps in stock price, no matter if the jump is up or down. So, if S turns out not to follow the Black and Scholes model as we postulated, but instead has a jump, or follows a Black and Scholes model but with higher volatility than we predicted (so that it oscillates in value more wildly than we predicted), this portfolio will be profitable.

However the equation $(\partial_y V)(x) = 0$ means that, for *infinitesimal* changes in value of S, the change of portfolio value is zero; this happens because the change $(\partial_y c)(x)$ in the call option price is offset by the change $\partial_y(c(x) + (y - x)\partial_x c(x))(x)$ in the value of the hedging portfolio. This strongly suggests that formula $G_t = \partial_x c(t, S_t)$ for the replicating strategy G is indeed correct.

If instead the stock does follow the Black and Scholes model with the correct σ, and $-\partial_x c(t, S_t)$ is indeed the hedging strategy, our portfolio is worth 0 at all times. How can this happen? After all, the change in portfolio value should be 0 only for *infinitesimal* changes in the value of the underlying, so that as time increases and the value of S changes from x to y the portfolio value should increase from 0 to $V(y) > 0$, because of this portfolio is long gamma (i.e. $\partial^2_{y^2} V > 0$). Moreover, as time increase the value of the bank account increases, since we deposited in it the positive amount $M > 0$. The reason why this does not happen (i.e. the value of V is always the constant 0) is that $\Theta = \partial_t c < 0$, so the increase in value of the portfolio due to the change in value of the underlying and the increase in value of the bank account is cancelled out by the decrease in value of the call option.