Department of Econometrics and Operations Research

Tilburg School of Economics and Management

Tilburg University

Introduction to Mathematical Finance

J.M. Schumacher

spring semester 2016

Contents

Chapter 1

Financial markets and institutions

1.1 The financial economy

English-Dutch vocabulary for Section 1.1

bond obligatie
cashflow kasstroom, ontvangen of gedane betaling
counterparty wederpartij, persoon of bedrijf waarmee een contract wordt aangegaan
demand vraag
due diligence gepaste waakzaamheid
exchange beurs
market value marktwaarde
over-the-counter contract maatcontract
real economy reële economie
solvency solvabiliteit
solvent solvabel, in staat aan de financiële verplichtingen te voldoen
stock aandeel
supply aanbod
utility nut
utility value nutswaarde

People need to be able to get loans for instance to buy a house, to put money away for later use into savings accounts that provide interest, and to buy insurance against the financial consequences of accidents. Companies need to fund their activities by a variety of instruments including stock, bonds, and loans from banks. If such facilities would not be available, economic life would come to a standstill. The role of financial markets and institutions is to provide these facilities. They make it possible for people and companies to move consumption forward or backward in time, and to modify exposure to risks. In other words, they allow redistribution of cashflows both *over time* and *across events*.

To understand the role of financial markets, it is necessary to distinguish between *market value* and *utility value*. In a fair trade between two parties, the market value of one side of the contract is equal to the market value of the other side of the contract. While the net market value of a fair contract is therefore zero, the contract can be utility improving for both parties. One side of the contract can be a good deal for one party, while the other side is a good deal for the counterparty as well, simply because different parties are in different positions, have different objectives, and are exposed to different risks. The economic significance of financial markets, and of dealers in the market, is to facilitate these trades which are beneficial to all parties involved.

Financial contracts may be constructed in a private agreement between two parties (*over-the-counter contract*) or they may be traded at an exchange. Over-the-counter contracts are fully adjustable while exchange-traded contracts must follow a limited number of standardized forms. While the standardization reduces flexibility, it does help to make contracts comparable so that parties can be more easily sure to pay a price that is reasonable given current market conditions. In many financial markets, over-the-counter and exchange-traded contracts are both used extensively.

Demand and supply are influenced by changing circumstances and by the arrival of new information. As a result, prices fluctuate. The existence of financial markets therefore does not only allow trading for economic reasons, as described above, but also *speculation*. Anyone who is able to predict the movement of prices has access to unlimited wealth; this fact will probably never lose its fascination to mankind. The *efficient market hypothesis* states that it is not possible to generate gains systematically by buying and selling in the market, since all available information is already taken into into account in current market prices, and therefore the movement of prices is determined by unpredictable future events. Although the validity of this hypothesis is not generally accepted, it is safe to say that beating the market is not easy.

All parties that are active in financial markets, and in particular financial institutions, are subject to *regulation*. Two main issues that are important in regulation are the *solvency* of financial institutions such as banks and insurance companies, and the *prevention of fraudulent behavior* by individuals who operate in the market. A well-known example of fraud in financial markets is the *Ponzi scheme*, which is based on paying high returns to investors using the money that is brought in by new investors. The scheme requires exponential growth of investments and is therefore difficult to sustain for a long time. The scheme is named after Charles Ponzi, an Italian immigrant to the US, who was able to maintain his swindle for about a year in 1919–1920. In addition to the surveillance of the authorities, market parties are expected to exert *due diligence* to affirm the trustworthiness of their counterparties.

1.2 Products and markets

A *financial product* is any contract that entitles or obliges the parties involved to receive or pay certain cashflows that are defined either in advance or in terms of information that will have become available at the time at which the cashflow takes place. For instance, when Alice borrows 100 euro from Bob and promises to give it back to him after one week, then this is a financial contract which involves two cashflows; both the sizes of the cashflows and the times at which they occur are fixed in advance. If Alice borrows 100 euro from Bob and promises to pay it back some time before the end of next month, then the sizes of the two cashflows are still fixed but the timing of the second cashflow is now decided, within

English-Dutch vocabulary for Section 1.2

asset actief (mv. activa), titel, iets van waarde
barrel vat
borrow from lenen van
call option kooprecht
collateral onderpand
contingent claim contract met variabele uitbetaling
convertible bond converteerbare obligatie
corporate bond bedrijfsobligatie
currency munteenheid
default verzaking (van verplichtingen), faillissement
dividend dividend (uitkering aan aandeelhouders)
entitle to recht geven op
exchange rate wisselkoers
forward contract leveringscontract
futures contract termijncontract
health insurance ziektekostenverzekering
interest rente
interest rate rentevoet
issue uitgeven
lend to lenen aan
maturity date afloopdatum
mortgage hypotheek
primary market primaire markt
put option verkooprecht
reimbursement vergoeding, terugbetaling
secondary market secundaire markt
share aandeel
stock aandeel
strike price uitoefenprijs
underlying onderliggende
volatility volatiliteit, beweeglijkheid

certain limits, by Alice. As another example, consider Carol's health insurance contract which obliges her to pay a yearly amount to the insurance company. The amount to be paid is specified in advance. When she needs medical care she sends the bills that she receives to the insurance company in order to receive reimbursement. These cashflows are not specified in advance. For a final example, suppose that David pays a certain amount for a contract that entitles him to receive, by July 1st of next year, the amount by which the market price of ten million barrels of oil at that time exceeds 800 million dollar. David may have good reasons to enter such a deal for instance when he runs an airline, because the contract gives him protection against high fuel prices. The amount that David pays to his counterparty is stated in the contract. The contract also states the formula for the amount that the counterparty will pay to David on the first of July of next year. The formula contains a variable that is not known at the time the contract is agreed between the parties, namely the price of oil on July 1 of next year, but this unknown will have become a known quantity at the time the payment will take place. A contract which determines a payment in terms of a variable whose value will be determined later is called a *contingent claim*.

Financial contracts can be agreed directly between parties, as in an over-the-counter contract. Financial institutions also offer certain standardized contracts, such as insurance contracts, savings accounts, and investment products. The term “financial product” can then be used, but actually there is no sharp distinction between “contracts” and “products”; the two terms will be used interchangeably below. Sometimes a financial contract may also be referred to as a *claim*, an *asset*, or a *security*.

A distinction can be made between *primary* and *secondary* markets. In primary markets, investors buy products directly from institutions that need to raise money, such as companies and governments. Such products may be resold to other investors in secondary markets. Reselling of financial products is quite standard in some cases, but highly unusual or even illegal in some other cases. Indeed it would be strange if Carol would sell her insurance contract to Eve, so that Eve would pay the premiums and would receive money from the insurance company when Carol gets sick. The insurance company on the other hand can resell part of the risk of its portfolio of insurance policies to another company. Reselling is a standard form of trading for many financial products.

There are many financial contracts that are familiar to practically everybody, such as savings accounts, insurance contracts, bank loans, and lottery tickets. Products that are well-known to anyone engaging in financial investment include stocks and bonds. A stock (or share) entitles the holder to dividend payments by the issuing company, and usually also give the holder some control of the company in the form of voting rights. The stream of dividends is uncertain so that it is not easy to determine the value of a stock; indeed the prices of stocks on the secondary market (the *stock market*) are noted for their fluctuations (volatility). Issuing shares is one way for a corporation to attract money from investors so that the firm's activities can be funded; another important way to raise money is by issuing bonds. Unlike stocks which entitle the holder to a stream of dividends with no termination date specified in advance, bonds are in fact loans which are paid back according to a predetermined schedule. This is not to say that bond holders do not run any risk; when the firm goes into default it may not keep its promise to pay back the loan. The interest on a corporate bond is therefore usually higher than on an otherwise comparable bond issued by an institution that is not likely to default, such as a national government or a municipality. There are also products that combine some of the characteristics of stocks and bonds, such as convertible bonds (bonds that may be converted to stocks). As in the case of stocks, there is a lively secondary market in which both corporate and government bonds are traded (the *bond market*).

The existence of highly liquid stock and bond markets has allowed the development of markets for all sorts of *derivative products*. These are products that define their payoff in terms of the market value of some underlying security which could be any asset that is actively traded. A standard example is the European call option on a stock, which gives the holder the right, but not the obligation, to purchase a certain quantity of the stock at a pre-specified date (the *maturity date*) at a pre-specified price (the *strike price*). The stock will be delivered by the institution that has sold ("written") the option. The investor who buys a call option can never lose more than the price of the option which is usually far below the price of the underlying stock, but the potential gain is unlimited. Of course the opposite holds for the institution that has written the option. A counterpart of the European call option is the European put option which gives the holder the right, but not the obligation, to sell a certain quantity of the stock at a pre-specified date at a pre-specified price. By buying a put option, the holder of the stock is protected, at least during a certain time, against downward movement of the stock price below the level of the strike price of the put. More generally, derivative products can be used to create a pattern of payoffs that will match the risk exposure of a particular person or company.

Derivatives may also written on certain events, such as a default event (the event that a particular company goes into default). Such derivatives can be useful to company A if it would suffer losses if company B would go bankrupt, for instance because the company B owes a debt to company A. Derivative contracts relating to commodity prices (oil, gold, wheat, etc.) are also very popular, since producers as well as consumers can use them to *hedge* against movements of prices that would be harmful to them.

A simple example of a derivative is the *forward* contract. Such contracts have existed since ancient times. In their basic form they are simply agreements between a producer and a consumer, stating that at a certain time in the future the producer will deliver a certain amount of goods to the consumer at a price that is specified in the agreement. Given that the market price at the delivery date will in general be different from the price agreed in the forward contract, there will be an advantage to either one of the two parties involved, but it is not known *ex ante* (i.e., beforehand) which of the two parties this will be. The point of the contract is that both parties reduce their uncertainty.

Forward contracts are useful instruments in a situation where a buyer and a seller have already decided to transact with each other. A producer or consumer who doesn't yet want to commit to a specific counterparty may still obtain protection against the variability of prices by buying a standardized *futures contract* at the *futures market*. Futures contracts are similar to forward contracts, but the number of different types of contracts is limited to facilitate trading. Moreover, settlement usually takes place in cash, rather than by delivery of the goods. This means that the holder of a futures contract receives the difference between the market price at the delivery date and the price that is stated in the contract, if this difference is positive; if on the other hand the difference is negative, then the contract holder is obliged to pay. Cash settlement has as a side effect that it facilitates the use of futures contracts for speculative purposes.

New markets have developed in recent years as a result of processes of *securitization*. Securitization means that financial contracts for which traditionally no secondary markets existed are turned into securities that can be traded. This is by no means a new idea; indeed, bonds are essentially securitized loans, and they have existed for hundreds of years. However the idea of securitization has been applied to new investment categories such as for instance mortgages. To securitize mortgages, they are packed together into *pools* which then are usually divided into *tranches* (also called *layers*). So, for instance, it would be possible to buy the claim to the first twenty percent of mortgage payments made by a group of home owners in a certain region and in a certain credit status category; "first" here means that these payments will be honored even if eighty percent of the expected mortgage payments are in fact not made by the home owners. When the home owners involved are of a reasonable credit status, the upper tranche is a rather safe investment, similar to a government bond, so that it would carry a comparable interest rate. There is also a second tranche which is somewhat less safe and which therefore has a somewhat higher rate attached to it to compensate for the default risk. The third tranche has an even higher rate attached to it and carries even more default risk, and so on. Such products are called *mortgage-backed securities*. They form a particular type of *collateralized debt obligations* (CDOs). Some remarks on the pricing of CDOs will be made in Section 5.2 below.

1.3 Financial institutions

English-Dutch vocabulary for Section 1.3

acquisition overname
ask price laatprijs
bid price biedprijs
casualty insurance ongevallenverzekering
commercial bank nutsbank, consumentenbank
credit line kredietlijn, rekening waarop de client rood kan staan
disability insurance arbeidsongeschiktheidsverzekering
health insurance ziektekostenverzekering
investment bank zakenbank
liability insurance aansprakelijkheidsverzekering
life insurance levensverzekering
management fee beheervergoeding
merger fusie
monetary policy monetair beleid
mutual fund beleggingsfonds
property insurance opstalverzekering
property-casualty insurance schadeverzekering
reinsurance herverzekering
retail bank consumentenbank
securitization securitisatie, verhandelbaarmaking
supervision toezicht

Most people prefer to keep their money in a bank account, rather than to put it under a pillow. *Commercial banks* (also called *retail banks*) are in the business of providing financial services to individual customers. They administer accounts, pay interest on savings, and provide loans to individuals and small businesses.

Large companies work with banks as well. All major firms have close relationships with banks that provide services such as

- a *credit line* which allows the company to borrow from the bank, up to a certain limit, so that there is a buffer for temporary imbalances between incoming and outgoing cashflows
- facilitating the sale of company shares and bonds to individuals as well as to institutional investors
- business advice, for instance concerning mergers and acquisitions.

Banks that engage in such activities are called *investment banks*.

The role of the *central bank* is to *supervise* financial institutions, to act as a *lender of last resort*, and to implement *monetary policy*. The aim of supervision is to ensure that financial institutions are *solvent*, which means that they will be able to meet their obligations also in times of hardship. Banks are for this reason required to keep certain reserves on an account with the central bank. Not only banks but also other financial institutions such as insurance companies and pension funds are subject to central bank supervision.

An important instrument of monetary policy is the interest rate that the central bank charges for loans. Loans from the central bank are always short-term loans, and therefore the rate set by the central bank has mainly an effect on short-term interest rates. Monetary policy aims to strike a balance between the possibly conflicting objectives of stimulating the economy and keeping inflation at a low (but not negative) level.

Insurance companies are financial institutions that are familiar to everyone. Well-known types of insurance are: health insurance; life insurance; disability insurance; liability insurance; property and casualty insurance. There are companies that specialize in one of these (*monoline* companies) and companies that have multiple lines of business (*multiline* companies). Some insurance companies have merged with banks to form conglomerates that offer a wide range of financial services. Sometimes insurance companies need to transfer some of the risk they assume, and for this purpose there are *reinsurance companies* which will cover, at a certain price of course, for instance the cost of claims in a certain line of business that exceeds a certain level (*stop-loss reinsurance*). As an alternative, insurance companies may attempt to shed excessive risk by bundling part of their business in products that may be sold to investors; this is called *securitization*.

Many people invest indirectly, for instance through a *mutual fund*. Some mutual funds do little more than following as closely as possible a certain benchmark such as the Dow Jones index, and they charge relatively low management fees. There are also funds that claim to obtain better results by the application of advanced strategies, and of course such funds charge higher fees. In particular *hedge funds* advertize themselves as being able to make money even in times when the stock markets go down. Whether such claims can be maintained is a much-debated issue; it appears that many funds have a hard time justifying high fees on the basis of actual investment results. Mutual funds and hedge funds are important players on the financial markets, together with other institutional investors such as pension funds.

1.4 Liquidity, efficiency, and rationality

Assets can be more or less *liquid*. A liquid asset is one for which there is an active market. It is easy to buy or sell such assets, and the price at which you can buy is close to the price at which you can sell. Examples of liquid assets include stocks of large companies, government bonds and so on; also many derivative contracts such as stock options, interest

English-Dutch vocabulary for Section 1.4

annuity annuïteit, lijfrente
bubble zeepbel
floating interest rate variabele rente
liquid liquide, veel verhandeld
liquidity liquiditeit

rate swaps, and oil futures are very liquid.[1] Examples of illiquid assets can be found in privately-owned companies, large real estate objects, and collector's items in exotic art. There are also illiquid liabilities: annuity providers such as pension funds would probably like to unload the longevity risk they are confronted with (the risk that the recipients of annuity payments will live longer than expected, so that more needs to be paid), but there is not much of a market for such risk. In liquid markets, the prices of assets may go up and down, but at any given moment in time the price is quite clearly defined, with only a very small difference between bid and ask price. For illiquid assets on the other hand, it is hard to quote a "market price", because of the lack of a market for such assets. Since investors are reluctant to buy illiquid assets, the price at which such an asset is traded may be lower than what the price of a comparable liquid asset would be; the difference is called the *liquidity premium*.

The main role of *exchanges* is to improve liquidity. Without the presence of an organized exchange, buyers and sellers would have a hard time finding each other, and trading would be a laborious affair. Stock exchanges have existed for more than four centuries, and exchanges for many other financial products have been created. The exchanges do not just provide a place for buyers and sellers to meet; in many cases they actively provide liquidity through so called *market makers*. Anyone who is interested in transacting some of the assets traded at the exchange can ask a market maker to provide a buying price (*bid price*) and a selling price (*ask price*); the market maker is obliged to give this information without knowing whether the client's intention is to buy or to sell. The client may then either sell to or buy from the market maker at the indicated price. Competition between different market makers at the same exchange helps to keep the *bid-ask spread* small.

A market is said to be *efficient* if all relevant information is available to and adequately processed by all market participants. In an inefficient market, some traders have an information advantage. Based on that advantage, they can make systematic profits. In many countries, extensive legal measures have been taken to support the efficiency of financial markets. For instance, companies are obliged by law to release immediately any information that may influence the price of the company stock. Company executives who trade on the basis of price-sensitive information that has not yet been made public are punishable by law, and they are not allowed to share such information with anyone. Since there is

[1] An *interest rate swap* is a contract in which a fixed interest rate payment is exchanged for a floating interest rate payment. An *oil future* is a contract to buy a given amount of oil of a given quality at a given time in the future.

no precise definition of “relevant information” nor of “adequate processing,” the notion of market efficiency is not defined very sharply. For instance, in modern financial markets some traders are able to buy and sell on a millisecond time scale, and are therefore able to process information faster than other traders; one may debate whether this constitutes a market inefficiency or not.

In the history of financial markets, *bubbles* are a well-known and frequently highlighted phenomenon. The term “bubble” is used for a period in which the prices of certain assets rise more quickly than would be justified by objective economic considerations. Prices are driven up by *speculation*. Even if you buy an asset for more than it is actually worth, you can still make a profit as long as the price goes up even further. This is a motivation for speculators to buy as long as the price is rising. Because there are people who buy, the price goes up even further. The process continues until at some point there are not sufficiently many speculators anymore who believe that the price will go up even higher. At that point the bubble bursts, there is a sharp decline in the asset price, and the last ones who bought are confronted with heavy losses. There are many historical examples of bubbles in financial markets, starting with the “tulip mania” in the 17th century Dutch Republic to the dotcom bubble at the end of the 20th century and the U.S. housing bubble in the early 21st century. It should be said though that in all cases there were some underlying economic reasons that could explain at least part of the price increases, and that it is easier to recognize a bubble after it has burst than before.

The appearance of bubbles in financial markets is often quoted as proof that investors show *irrational* behavior. However, it is not irrational to buy an asset at more than what you think is its true value, if you believe that you can sell it again at an even higher price. Theoretical models show that bubbles can arise in markets in which all participants are rational, but in which there are participants who *believe* (mistakenly) that there are participants who are not rational. Even if in a market everybody is rational and everybody believes that everybody is rational, but there is somebody who thinks (mistakenly) that some participants believe (mistakenly) that there are some other participants who are irrational, then this may still be enough to create a bubble. So it is not just irrationality that may create a bubble, but also just suspicion of irrationality, or even only suspicion of suspicion of irrationality, and so on. Given also the fact that the notion of “true value” is quite elusive, the actual presence of irrationality in the market may not be so easily demonstrated as is sometimes thought.

In the present course, it will be assumed that markets are efficient and are populated by rational investors with no speculative intentions. In the discussion of derivative assets, it will also be assumed that the basic (underlying) assets are completely liquid and that there is in fact no difference between bid and ask price. These assumptions are convenient for mathematical modeling, and they are valid to good approximation in highly developed markets during normal times of operation. However, in less developed markets, and also in highly developed markets either in times of crisis or in times of exuberance, the deviation between these assumptions and reality may not be negligible.

1.5 Outline of the course

English-Dutch vocabulary for Section 1.5

balance afweging	**mathematical finance** financiële wiskunde
hedge indekken, afdekken	**portfolio** portefeuille
liability financiële verplichting, te verrichten betaling	**preference** voorkeur
	return rendement
market consistent marktconsistent	**valuation** waardering

The main topics in mathematical finance are the following.

- Valuation: what is a reasonable price to pay for a financial contract.
- Risk Management: how to hedge against price risk, currency risk, interest rate risk, and other risks.
- Portfolio Management: optimization of investment strategies, taking into account existing *liabilities* and a proper *balance between risk and return*.
- Financial Engineering: *design* of financial products.

A large part of these notes is dedicated to valuation. Both valuation of *deterministic* and of *stochastic* cashflows will be addressed. We start with *fundamental valuation*, which is based on assumptions concerning the preferences of market participants. The main focus however will be on *market-consistent valuation* in which prices of contracts are determined on the basis of observed prices of related contracts. It turns out that market-consistent valuation and risk management are closely related. The final chapter is concerned with portfolio management and in particular devotes attention to an optimization technique that is built on ideas that are also used for option pricing.

A mathematical model of a financial market is essentially a specification of the cashflows that are available for trading. In these notes, three different types of specifications will be used. The first type consists of *multiperiod markets without uncertainty*, the second type refers to *single-period markets with uncertainty*, and finally the third type covers *multiperiod markets with uncertainty*. These types are graphically illustrated in Fig. 1.1. The first two types are mathematically similar, in the sense that the collection of cashflows available to agents (i.e. the available assets) can be described in terms of a matrix; rows correspond to different assets, and columns correspond to either different *time points* (in the case of a multiperiod market without uncertainty), or to different *events* that may take place (in the case of a single-period market with uncertainty). The multiperiod model with uncertainty however must be described in a different way. More advanced models in mathematical finance are based on stochastic processes that define the random evolution of continuous variables in continuous time, but such models are outside the scope of these course notes.

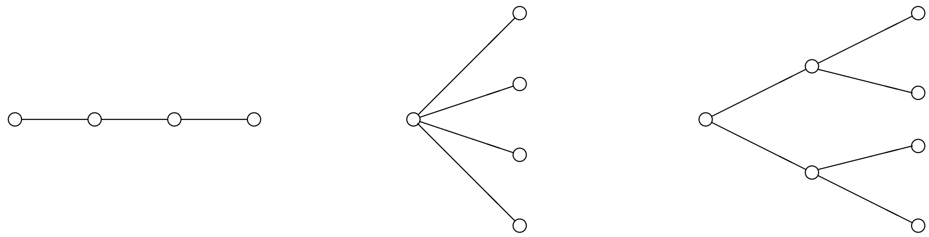

Figure 1.1: Left: multiperiod market without uncertainty. Middle: single-period market with uncertainty. Right: multiperiod market with uncertainty.

1.6 Exercises

1. The valuation of bonds depends on the credit status of the institution that issues the bond; if there is a non-negligible probability that the issuer (which can be a corporation, a local government, or a national government) will default on its liabilities, so that bond holders do not get their money back, then the existence of that probability will have an effect on bond prices. The credit status of corporations and governments is monitored by rating agencies who place bond issuers in various categories of reliability such as AAA (very trustworthy), AA (somewhat less reliable but still quite safe), A, BBB, and so on down to C and finally D (default). In a fully efficient market, do you expect that bond prices will be affected when a rating agency moves an issuer to a different rating category?

Chapter 2

Portfolio choice

English-Dutch vocabulary for Section 2.0	
portfolio portefeuille	**utility function** nutsfunctie
utility nut	

2.1 Decision under uncertainty

Suppose that you find yourself in a situation in which you have to make an investment decision. Several investment opportunities are available, each with their own risk profile. You have to decide how much of your available budget to invest in each of these. What is the best choice? The answer may depend on your degree of risk aversion; if taking risk makes you lie awake at night, then probably you should choose investment that have a low risk profile. The answer also depends on how attractive the opportunities are, at least in your eyes; if you think that the probability is high of getting a very nice return from a particular investment, then you may be tempted by that.

The situation described in the previous paragraph is one of *decision under uncertainty*. Problems in this area are as old as the world; mathematical approaches have been developed since the 18th century. A standard formalization, originally proposed by Daniel Bernoulli in 1738 and still popular today, is based on the notion of "expected utility". The theory exists in single-period as well as in multiperiod and continuous-time versions. Within this formalization, the decision that should be taken is the one that achieves the highest expected value of the utility of the payoff. To apply this framework, the following elements are needed.

- First: a utility function. Typical assumptions are that the utility function is increasing and concave. It is also often assumed that the utility function is twice differentiable.
- Secondly: an assignment of probabilities to possible outcomes. In some situations, for instance when placing best at the roulette table, it is possible to make precise

statement about the probabilities of outcomes; one then speaks of “objective” probabilities. In most real-life economic situations, however, probabilities are subjective to a smaller or larger degree, and the choice of probabilities should then be considered as a characteristic of the individual decision maker, just as the utility function is.

- Thirdly: a specification of the set of feasible decisions. In other words, the constraints under which the decision is taken should be specified. There is always a budget constraint, but there may be other constraints as well. For instance, additional constraints may result from indivisibility (you cannot buy arbitrary fractions of a certain asset, but only chunks of a certain minimum size), short-selling restrictions (you are not able to hold negative positions), or market incompleteness (you can only invest in a limited number of assets, which are not enough to create arbitrary payoff profiles).

The most frequently used utility functions are the following:

- constant absolute risk aversion (CARA; exponential utility)

$$u(x) = -ae^{-x/a} \qquad (\alpha > 0) \tag{2.1}$$

- constant relative risk aversion (CRRA; power utility, with log utility as a special case; only for $x > 0$)

$$u(x) = \frac{x^{1-\gamma} - 1}{1-\gamma} \qquad (\gamma \geq 0,\ \gamma \neq 1) \tag{2.2}$$

$$u(x) = \ln x \quad \text{(limit of the above as } \gamma \to 1\text{).} \tag{2.3}$$

The framework of expected utility has been extensively used within economic theory. However it has also often been criticized for not providing a good model of actual human behavior. Indeed it is true that individuals typically find it difficult to express their preferences in terms of a utility function. Many alternative models have been proposed, which may allow more direct interpretation or which may provide a more accurate description of preferences. Examples of alternative models include the following.

- *Mean-variance preferences*, expressed by a function of the form

$$E(V_T) - \gamma\,\mathrm{std}(V_T)$$

with $\gamma > 0$. The interpretation of this specification is quite direct, and the degree of risk aversion is expressed in terms of a single parameter γ. A disadvantage of the use of variance as an indicator of risk is that this quantity does not distinguish between positive and negative deviations. Expression of risk aversion by a single parameter is also available within the expected utility framework by restriction to a specific class such as CARA or CRRA.

- *Robust expected utility* given by

$$\min_{i=1,\ldots,n} E^{P_i}(u(V_T))$$

where $P_1, \ldots, P_n$ is a collection of probability measures that are considered "reasonable". This specification takes into account that in many real-life situations it is hard to be precise about the probabilities of future events, and people may take this form of uncertainty (also known as *ambiguity*) into account when taking decisions. By taking a minimum across possible assignments of probabilities, the framework of robust expected utility takes a worst-case approach. To apply the framework, one has to provide a utility function as well as a set of "reasonable" probability measures.

- Addition of *side constraints*, such as the condition that the probability of portfolio value to fall below a given level should not be larger than 1%. Some side constraints, but not all, can be accommodated within the expected utility framework; for instance the condition that capital should not fall below a certain minimum level can be expressed by a utility function that takes the value $-\infty$ for values below the specified minimum level.

In view of the many approaches that are possible, one should always be careful to specify what is meant when the word "optimal" is used. In as far as specification of preferences will be needed in these course notes, we will remain within the expected utility framework.

2.2 Single-period discrete-state model

Here we consider a single-period situation in which uncertainty is represented by a finite number of possible outcomes, numbered from 1 to n. The outcomes are called "states of the world", or simply "states". At time 1 (end of period), one of the states will be revealed as the actual state, but this information is not available at time 0 (beginning of period). Investment opportunities are given by *assets*. In the single-period context, an asset is characterized on the one hand by its *price* (a single number, representing the per-unit price one has to pay at time 0 to buy the asset) and on the other hand by its *payoff* (a vector of length n, in which the entry with index i represents the monetary payoff that will be produced at time 1 by one unit of the asset if state i materializes). We will consider the decision problem under the assumption that only the budget constraint needs to be taken into account. This means for instance that one can form portfolios with 3.14159 units of one asset and -6.5 units of another, and that the payoff vectors of the available assets span the n-dimensional space. The price of a portfolio will always be assumed to be determined linearly; that is to say, the price of the portfolio consisting of a_1 units of asset 1, a_2 units of asset 2, and so on, is given by a_1 times the per-unit price of asset 1 plus a_2 times the per-unit price of asset 2, and so on.

Under the spanning assumption and the linear pricing rule, it is no restriction of generality to assume that the available assets are actually the ones whose payoff vectors are the unit vectors of the n-dimensional vector space. Such assets are called *state contracts*. The corresponding prices are called *state prices* and will be denoted by π_i. In other words, π_i is the price you need to pay at time 0 (measured in units of currency) to buy an asset that will pay one unit of currency at time 1 if state i is drawn, and nothing if another state is realized.[1]

In this formulation, the time at which payment is made is different the time at which the payoff is received. The payment is expressed in time-0 units of currency, while the payoff is expressed in time-1 units of currency. Therefore in general the price of a contract that pays 1 with certainty is not equal to 1, but a different number, which we write as $1/(1+r)$. The quantity r is called the *simply compounded interest rate*. Since the payoff 1 in all states of the world is achieved by a portfolio that contains one unit of all the state contracts, the following relation holds:

$$\frac{1}{1+r} = \sum_{i=1}^{n} \pi_i. \tag{2.4}$$

It is often convenient to work with *normalized state prices*. These are denoted by q_i and defined by

$$q_i = (1+r)\pi_i = \frac{\pi_i}{\sum_{i=1}^{n} \pi_i}\,. \tag{2.5}$$

Normalized prices can be viewed as prices expressed in terms of time-1 units of currency. It will be assumed that state prices are positive, as is not unreasonable, and as will also be further motivated in a later chapter. The normalized state prices are then positive as well. Note that the sum of the q_i's is by definition equal to 1.

In the situation described above, consider now an investor whose available budget is given by W, who has a utility function given by $u(x)$, and who assigns probabilities $p_1, \ldots, p_n$ to the outcomes $1, \ldots, n$. Any portfolio that the investor can choose can be described as a linear combination of state contracts. The investor's decision variables can therefore be taken to be the numbers $x_1, \ldots, x_n$, where x_i is the amount invested in state contract i. The budget constraint is given by

$$\sum_{i=1}^{n} \pi_i x_i = W. \tag{2.6}$$

This can also be written, in terms of the normalized prices, as

$$\sum_{i=1}^{n} q_i x_i = (1+r)W. \tag{2.7}$$

[1] In the context of betting, the price is usually represented in another way: you pay one unit of currency, and you receive α units of currency if the event on which you bet is realized. The corresponding state price as defined here is given by $\pi = 1/\alpha$.

The payoff obtained from the portfolio given by the holdings vector $(x_1, \ldots, x_n)$ is a random variable X that takes the value x_i with probability p_i. The corresponding expected utility is given by

$$E[u(X)] = \sum_{i=1}^{n} p_i u(x_i). \tag{2.8}$$

The investor's optimization problem can now be formulated as:

$$\begin{aligned} &\text{maximize} \quad && \sum_{i=1}^{n} p_i u(x_i) \\ &\text{subject to} \quad && \sum_{i=1}^{n} q_i x_i = (1+r)W. \end{aligned}$$

The problem as stated above is suited for an application of the Lagrange method for optimization subject to equality constraints. We work under the usual assumptions that the utility function $u(x)$ is strictly increasing, strictly concave, and twice differentiable, so that the first derivative $u'(x)$ is positive and the second derivative $u''(x)$ is negative. There is a single constraint, namely the budget constraint (2.7); the corresponding Lagrange multiplier will be denoted by λ. The Lagrangian function for the problem is

$$L(x_1, \ldots, x_n, \lambda) = \sum_{i=1}^{n} p_i u(x_i) - \lambda\Big(\sum_{i=1}^{n} q_i x_i - (1+r)W\Big).$$

Setting the partial derivatives of this function with respect to the decision variables x_i equal to zero leads to the equations

$$p_i u'(x_i) = \lambda\, q_i \qquad (i = 1, \ldots, n). \tag{2.9}$$

The derivative $u'(x)$, which is also known as the marginal utility, is by assumption a strictly decreasing function. Therefore it has an inverse function, which will be denoted by $v(\cdot)$; in other words, the function $v(\cdot)$ is such that $v(u'(x)) = x$ for all x. Table 2.1 shows the inverse marginal utilities for some of the most frequently used utility functions. It also shows the corresponding risk aversion functions $R(x)$, which are defined by $R(x) = -u''(x)/u'(x)$. In terms of the inverse marginal utility $v(\cdot)$, the solution to the equation above can be written as

$$x_i = v\Big(\lambda \frac{q_i}{p_i}\Big). \tag{2.10}$$

The optimal decisions x_i are now parametrized in terms of a single variable, namely the Lagrange multiplier λ. This variable is determined from the corresponding equality constraint:

$$\sum_{i=1}^{n} q_i v\Big(\lambda \frac{q_i}{p_i}\Big) = (1+r)W. \tag{2.11}$$

One might also say that the equation (2.10) gives all optimal decisions for various possible available budgets, and that the required budget corresponding to each of these decisions is given by (2.11).

name	domain	$u(x)$	parameter restrictions	$u'(x)$	$v(x)$	$R(x)$
log utility	$(0,\infty)$	$\ln x$	not applicable	$1/x$	$1/x$	$1/x$
power utility	$(0,\infty)$	$\frac{x^{1-\gamma}}{1-\gamma}$	$\gamma > 0, \gamma \neq 1$	$x^{-\gamma}$	$x^{-1/\gamma}$	γ/x
exponential utility	$(-\infty,\infty)$	$-\tau e^{-x/\tau}$	$\tau > 0$	$e^{-x/\tau}$	$-\tau \ln x$	$1/\tau$

Table 2.1: Popular utility functions

2.3 Alternative representations

The form of the solution allows some further interpretation. From the expression (2.10), it is seen that the amount of money to be invested in the i-th state contract depends on the quotient q_i/p_i. This quotient can be viewed as being similar to a price/performance ratio, since q_i is the normalized price of one unit of state contract i, and p_i is its expected payoff. Because both the q_i's and the p_i's sum to 1, not all of the q_i's can be larger than the corresponding p_i's, or vice versa; in other words, if there is an index i such that $q_i/p_i > 1$, then there is also an index j such that $q_j/p_j < 1$. States for which the ratio q_i/p_i is larger than 1 might be called "overvalued" (at least in the eyes of the investor who assigned the probabilities p_i), and states for which $q_i/p_i < 1$ might be called "undervalued". Because the marginal utility $u'(\cdot)$ is a decreasing function, its inverse function $v(\cdot)$ is decreasing as well. From (2.9) it follows that λ must be positive. Combining these two observations, it is seen from (2.10) that, irrespective of the available budget, in the optimal solution we always have

$$x_i > x_j \quad \text{if and only if} \quad q_i/p_i < q_j/p_j. \tag{2.12}$$

In other words, higher stakes are placed on outcomes that are more strongly undervalued, which is indeed sensible.

A more quantitative description of the optimal solution can be obtained as follows. Without loss of generality, we can assume that the states are numbered in order of attractiveness, so that

$$q_1/p_1 \leq q_2/p_2 \leq \cdots \leq q_n/p_n.$$

In other words, the most undervalued state is numbered 1, and the most overvalued state gets number n. Then we already know from (2.12) that the optimal solution is such that

$$x_1 \geq x_2 \geq \cdots \geq x_n.$$

From the optimality condition (2.9), it follows that for all i and j the following relation holds:

$$\frac{u'(x_i)}{u'(x_j)} = \frac{q_i/p_i}{q_j/p_j}\,.$$

Taking logarithms, this may also be written as

$$\ln u'(x_i) - \ln u'(x_j) = \ln \frac{q_i}{p_i} - \ln \frac{q_j}{p_j}\,. \tag{2.13}$$

Now note that the coefficient of absolute risk aversion

$$R(x) := -\frac{u''(x)}{u'(x)}$$

is actually equal to minus the derivative of $\ln u'(x)$. This means that the left hand side of (2.13) can be written as an integral:

$$\ln u'(x_i) - \ln u'(x_j) = \int_{x_j}^{x_i} R(x)\,dx.$$

Consequently, any optimal solution must satisfy

$$\int_{x_j}^{x_i} R(x)\,dx = \ln \frac{q_i}{p_i} - \ln \frac{q_j}{p_j} \quad \text{for all } i \text{ and } j. \tag{2.14}$$

Optimal solutions can therefore be constructed as follows, from a given set of p_i's and q_i's and a given risk aversion function $R(x)$. Start with a guess x_1 for the amount of units to be bought of the most favorable state contract. Then choose $x_2 \leq x_1$ such that

$$\int_{x_2}^{x_1} R(x)\,dx = \ln \frac{q_2}{p_2} - \ln \frac{q_1}{p_1}\,.$$

Then choose x_3 such that the integral of the risk aversion coefficient $R(x)$ from x_3 to x_2 is equal to $\ln q_3/p_3 - \ln q_2/p_2$, and so on. After all x_i's have been constructed, check whether $\sum_{i=1}^{n} q_i x_i$ is smaller than or larger than $(1+r)W$. If the sum is smaller, then x_1 needs to be increased; if it is larger, than x_1 must be decreased. This procedure is not particularly efficient from a purely computational point of view, but it may help to get an intuitive grasp on the interaction of on the one hand the risk aversion function $R(x)$, and on the other hand the ratios q_i/p_i. Here are some observations that can be made.

- The differences between the x_i's will tend to get smaller when the investor becomes more risk averse. From a mathematical point of view, this is a consequence of the fact that, in order to keep the value of an integral the same when the integrand becomes larger, the domain of the integration must become smaller. It also makes sense intuitively: a more risk averse investor will opt for a smaller variability across possible outcomes.

- Other things remaining equal, the differences between the x_i's will tend to get larger when the ratios q_i/p_i are spread out more widely. This is seen from equation (2.14). Again, the statement makes intuitive sense: an investor who perceives strong undervaluation or overvaluation of some states will be enticed to follow a relatively aggressive strategy. At the other extreme, when all q_i's are equal to the corresponding p_i's, so that $q_i/p_i = 1$ for all i, then all x_i's are the same; in other words, the investor takes a riskless position.

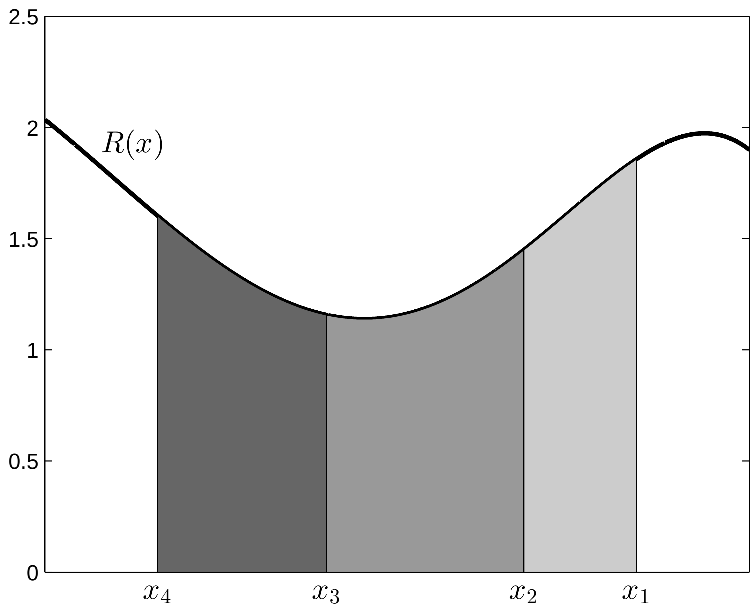

Figure 2.1: Optimality condition. The bold curve shows the coefficient of relative risk aversion as a function of possible payoffs. The surface area under the curve between the points x_2 and x_1 is equal to the quantity $\ln \frac{q_2}{p_2} - \ln \frac{q_1}{p_1}$, and likewise for the other shaded areas.

- When the budget is increased, then all x_i's become larger. Indeed, if x_1 is moved to the right, then x_2 must also move to the right to keep the same value for the integral $\int_{x_2}^{x_1} R(x)\, dx$. As a consequence, x_3 must likewise move to the right, and so on. This shows that, in the optimal solution for a given risk aversion function $A(x)$ and ratios q_i/p_i, the x_i's must move in unison. This is seen from the expression (2.10) as well.

2.4 Exercises

1. The utility functions $u(x)$ and $\tilde{u}(x)$ are said to be *equivalent* if there exist a positive constant a and a constant b such that

$$\tilde{u}(x) = au(x) + b$$

for all x. (The relationship implies that $E[\tilde{u}(X)] = aE[u(X)] + b$ for all random variables X, so that optimizing $E[\tilde{u}(X)]$ over a given set of random variables X leads to the same solution as optimizing $E[u(X)]$.) Show that two utility functions $u(x)$ and $\tilde{u}(x)$ are equivalent if and only if the corresponding risk aversion functions $R(x)$ and $\tilde{R}(x)$ are the same.

2. Verify that, for any fixed $x > 0$, we have

$$\lim_{\gamma \to 1} \frac{x^{1-\gamma} - 1}{1 - \gamma} = \ln x.$$

3. Consider an agent whose preferences are given by expected utility, with differentiable and strictly concave utility function $u(x)$ (i.e. the derivative $u'(x)$ exists and is strictly decreasing for all x). Assume that prices of state contracts are given by π_j. Supposed that a proposed allocation $(x^1, \ldots, x^n)$ for the agent does *not* satisfy the proportionality condition

$$p_i u'(x_i) \propto \pi_i.$$

Show one way in which the agent can achieve a higher level of expected utility by trading, while respecting the budget constraint. [Hint: if the proportionality condition is not satisfied, then there are states, say 1 and 2, such that $p_1 u'(x_1) = c_1 \pi_1$ and $p_2 u'(x_2) = c_2 \pi_2$ with $c_1 \neq c_2$. Consider the expected utilities that can be achieved by changing the amounts of the two state contracts subject to the budget constraint.]

4. Consider a situation in which there are two possible states of the world, which have equal probability. Alice has a capital of 100 available, which she must divide between a state 1 contract (an asset which has a positive payoff if state 1 is realized, and nothing if state 2 is realized) and a state 2 contract (the other way around). The payoff of contract 1 in state 1 per invested unit is η times as large as the payoff of state 2 contract in state 2, with $\eta > 1$.

a. Suppose that Alice uses power utility with coefficient γ to determine her optimal choice of investments x_1 in the state 1 contract and x_2 in the state 2 contract. Show that her optimal choice satisfies

$$\frac{x_1}{x_2} = \eta^{1/\gamma}.$$

Note that, together with the budget constraint $x_1 + x_2 = 100$, the above rule is enough to determine the values of x_1 and x_2. What happens to the solution if γ tends to ∞? What happens if γ tends to 0?

b. Now suppose that Alice uses exponential utility, with risk tolerance parameter τ. Show that in this case her optimal choice is such that

$$x_1 - x_2 = \tau \ln \eta.$$

Again, together with the budget constraint, the above rule is enough to determine Alice's decision. What happens to the solution if the risk tolerance parameter τ tends to 0? What happens if τ tends to infinity?

5. **a**. Let a utility function be given of the form

$$u(x) = \frac{(x - x_0)^{1-\gamma}}{1 - \gamma} \tag{2.15}$$

with $\gamma > 0$; in other words, this is power utility with respect to the difference of the outcome x and the "subsistence value" x_0. Show that the corresponding absolute risk aversion function is of the form

$$R(x) = \frac{1}{ax + b} \tag{2.16}$$

for suitable constants a and b. [Since the graph of the function on the right hand side is a hyperbola, the utility function $u(x)$ defined by (2.15) is said to belong to the "HARA class" (hyperbolic absolute risk aversion).]

b. Show that exponential utility functions belong to the HARA class as well.

c. Show that, if a utility function belongs to the HARA class, then (up to multiplication by a positive constant and addition of a constant) it is either a shifted power utility as in part a., or a power utility. In other words, there are essentially no other utility functions in this class besides the ones that are already identified in parts a. and b. of this exercise.

6. **a**. Consider an economy with n possible future states of the world, which have probabilities $p_1, \ldots, p_n$ respectively; normalized state prices are given by $q_1, \ldots, q_n$, and the single-period interest rate is r. Write explicit formulas for the optimal portfolio for an investor with wealth W in each of the three cases identified in Table 2.1 (log utility, power utility, and exponential utility).

b. Suppose that Alice and Bob are both log utility investors, but Alice is twice as wealthy as Bob. How do their optimal portfolios relate to each other? Consider also the situation in which Alice and Bob are both power utility investors with the same risk aversion coefficient γ, and the situation in which they are both exponential utility investors with the same risk tolerance coefficient τ. How do the portfolios of Alice and Bob compare if they are both exponential utility investors, but their risk tolerance coefficients are different?

c. The *risk tolerance function* is defined by $T(x) := 1/R(x)$. Its derivative is called *cautiousness*. Investors are of HARA type (see exc. 5) if and only if their cautiousness function is a constant; this constant must be either positive or zero. Show that, for a given level of cautiousness, there exist two portfolios ("funds") such that the optimal portfolio of any HARA investor with that level of cautiousness is a linear combination of these two funds. [This is called "two-fund separation".]

7. Suppose that $u_A(x)$ and $u_B(x)$ are two utility functions, both strictly increasing, twice differentiable, and strictly concave, and that $R_A(x)$ and $R_B(x)$ are the corresponding risk aversion functions.

a. We first establish a mathematical fact that we be needed further below. If f is a strictly monotonic function, then the corresponding *inverse function*, denoted by f^{-1}, is the function defined by

$$x = f^{-1}(y) \quad \Leftrightarrow \quad f(x) = y.$$

In other words, $f^{-1}(y)$ is the function such that $f^{-1}(f(x)) = x$. Given that the function f

is differentiable, show that the derivative of its inverse function is given by

$$(f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))}\,.$$

b. Suppose that Alice and Bob have utility functions $u_A(x)$ and $u_B(x)$ respectively; both utility functions are strictly increasing, strictly concave,[2] and twice differentiable. Let the corresponding risk aversion functions be denoted by $R_A(x)$ and $R_B(x)$. We say that Alice is *strictly more risk averse* than Bob if

$$R_A(x) > R_B(x) \qquad \text{for all } x.$$

Prove that this relation holds if and only if there exists a strictly increasing and strictly concave function g such that $u_A(x) = g(u_B(x))$. [Hint: there is only one candidate, namely the function g defined by $g(z) = u_A(u_B^{-1}(z))$.]

c. Let X be a random variable that takes the value x_1 with probability p and the value x_2 with probability $1-p$; suppose that $x_2 > x_1$. Show that there exist a unique number x_A with $x_1 < x_A < x_2$ such that

$$u_A(x_A) = E[u_A(X)] = pu_A(x_1) + (1-p)u_A(x_2).$$

The number x_A is said to be Alice's *certainty equivalent* of the random variable X. In the same way one can defined Bob's certainty equivalent, indicated by x_B.

d. Show that, in the situation of part c., if Alice is strictly more risk averse than Bob, then

$$x_A < x_B. \tag{2.17}$$

[Hint: by redefining the utility scales, you can assume that $u_A(x_1) = u_B(x_1) = 0$, and $u_A(x_2) = u_B(x_2) = 1$. Use part b.] What is the economic meaning of the relation (2.17)?

8. **a**. Consider a model with two possible future states, which are labeled "u" (for "up") and "d" (for "down"). Investors have available a risky asset satisfying $S_\mathrm{u} = uS_0$, $S_\mathrm{d} = dS_0$, and a riskless asset satisfying $B_\mathrm{u} = B_\mathrm{d} = (1+r)B_0$.[3] Let the normalized state prices be denoted by q_u and q_d, and let p_u and p_d be the objective probabilities. An initial budget V_0 is given. Solve the optimization problem

$$\begin{aligned} &\text{maximize} \qquad p_\mathrm{u} u(V_\mathrm{u}) + p_\mathrm{d} u(V_\mathrm{d}) \\ &\text{subject to} \qquad q_\mathrm{u} V_\mathrm{u} + q_\mathrm{d} V_\mathrm{d} = (1+r)V_0 \end{aligned}$$

[2] A function $u(x)$ is said to be *strictly concave* if $u(\lambda x + (1-\lambda)y) > \lambda u(x) + (1-\lambda)u(y)$ for all x and y in the domain of u and for all λ such that $0 < \lambda < 1$. If the function u is twice differentiable, then an equivalent condition is that $u''(x) < 0$ for all x.

[3] The notation that is used here is common when the number of states is equal to two. The symbol u is also used for utility functions, and the symbol d for differentials: this should generate no confusion.

when $u(x)$ is a function of the form

$$u(x) = -\tau e^{-x/\tau}$$

(exponential utility).

b. Determine the corresponding optimal amount to be invested in risky assets.

c. Find the optimal amount to be invested in risky assets when the parameter values are as follows: $u = 1.2$, $d = 0.8$, $r = 0.04$, $p_{\text{u}} = 0.7$, $a = 10\,000$ euro.

9. Let a utility function be given by $u(x) = -\tau e^{-x/\tau}$ with $\tau > 0$ (exponential utility).

a. Let Z be a random variable that follows an $N(\mu, \sigma^2)$ distribution, and let x_0 be a given number. Given σ, determine the value that μ should have so that the following relation holds:

$$E\big(u(x_0 + Z)\big) = u(x_0).$$

b. Alice assumes that Bob uses an exponential utility function to evaluate investment opportunities, and she wants to know his parameter τ. She asks Bob the following question. "Suppose that a possible investment is presented to you whose payoff follows a normal distribution, with a one-in-two probability of generating a net profit of 5000 euro or more, and a one-in-six probability of generating a net loss of 5000 euro or more. Would you take it?" Given the result of part a., can you explain Alice's question? [This is an example of what is called *elicitation of preferences*.]

Chapter 3

Equilibrium

3.1 The notion of general equilibrium

English-Dutch vocabulary for Section 3.1

agent economisch agent; actor
equilibrium evenwicht
market clearing marktklaring; situatie waarin voor elk van de op de markt verhandelde goederen geldt dat de vraag gelijk is aan het aanbod
risk aversion risico-aversie, afkeer van risico

General equilibrium theory is concerned with markets in which several goods are traded. Supply and demand for each good depend on the market prices for all goods. If the prices are such that there is equilibrium between supply and demand for every good (*market clearing*), then we speak of "general equilibrium". The term "competitive equilibrium" is also used to emphasize that equilibrium is reached by adjustment of prices in a situation of free trading, rather than for instance by imposing trade limitations.

In order to use general equilibrium theory to determine prices, one has to specify in which way prices have an effect on supply and demand. Here one has to take into account direct effects (an increase of the price of coffee may reduce the demand for coffee) as well as indirect effects (an increase of the price of coffee may raise the demand for tea). An essential element of equilibrium models is therefore a specification of agents' preferences across "consumption bundles".

In general equilibrium theory, the objects that are being traded are often simply called *goods*. In examples that illustrate the theory, these goods are usually taken to be commodities, for instance coconuts. Other interpretations are possible, however. For instance, a "good" could be a cashflow that will take place at a certain time if at that time a given condition is satisfied. An example of such a good is a lottery ticket; the lucky number will be drawn at a preset time, and the condition that must be satisfied for you to win the jackpot is that the number drawn by the notary matches the number on your ticket. In this chapter we are concerned with the pricing of such "goods", which are also called *(financial) assets* or *contingent claims*. An asset is anything that can be used as a vehicle of investment, such as stocks, bonds, and so on. A contingent claim is a contract that entitles the holder to a possibly uncertain future payment. Contingent claims can be used as assets and, vice versa, assets can be viewed as contingent claims, so that there is no sharp distinction between these two terms.

Specially constructed contingent claims can be used by economic agents such as companies, governments, and individuals, to reduce the risks that they run as a result of their activities. For instance, airlines need to buy kerosine and are therefore naturally exposed to fluctuations of oil prices. An airline may therefore be interested in buying contracts that will pay a given amount of dollars for every dollar that the price of a barrel oil at a given date in the future exceeds a certain level. Such a contract provides protection against future increases of the price of kerosine. Of course there has to be a counterparty, who will be liable to make the payment in case the conditions of the contract are triggered. After the contract has been signed, part of the risk of oil price increases is borne by the seller; in other words, a *transfer of risk* has taken place. The seller will be willing to accept the risk if the price of the contract is sufficiently attractive. Given that the buyer and the seller have different exposures to risk, it may well happen that the contract is attractive to *both* of them from a utility point of view. Perhaps the institution that acts as the seller is itself in a position in which it makes more money when the oil price goes up, or it may be able to re-sell the risk to a third party who is in such a position. The latter possibility arises in particular when the seller is a financial institution. In this way, financial institutions play a role as mediators of *risk sharing* between agents in the real economy.

Example 3.1.1 Suppose that Alice runs an ice cream parlor and that Bob owns a cinema. As an asset, we can take a day's income. The income depends on the weather. In suitable units, the initial endowment of Alice is $(20, 5)$ and the initial endowment of Bob is $(10, 15)$, where the first component of both vectors represents "income on a sunny day" and the second component refers to "income on a rainy day".[1] Bob and Alice can agree to a contract in which Bob transfers some money to Alice on a rainy day, and Alice transfers some money to Bob when the weather is nice. By doing so, they both reduce their risks. If Alice and Bob are risk averse, the contract can be attractive to both of them, even though it is a zero-sum agreement in the sense that, on a sunny as well as on a rainy day, one agent's loss from the contract is the other agent's gain from it.

Potential buyers and sellers of risks can meet at financial markets. Since it may not be easy for buyers and sellers of specific risks to find each other, a large part of trading takes place through standardized contracts which are offered through intermediaries. Financial intermediation is one of the main services that are carried out by banks.

[1] As is seen here, the term "initial endowment" may well relate to payoffs that take place in the future, in contrast to what the term "initial" might suggest.

General equilibrium theory aims at predicting the outcomes of price negotiation processes that take place in the market. Prices are determined by demand and supply, which in turn are determined by the positions and preferences of agents. The theory is successful in providing a method to compute equilibrium prices, but for this it requires a full specification of the positions and preferences of all agents. To provide such a specification for the world's economy is in no way feasible. It is possible though to find equilibrium prices for simple artificial economies. By modeling these economies after selected features of the real world, one can then still try to use the theory to get insights that will be helpful in understanding what happens in actual financial markets. In this chapter we develop a few basic models and discuss the meaning of the equilibrium outcomes.

3.2 A review of general equilibrium mathematics

English-Dutch vocabulary for Section 3.2

allocation allocatie, toewijzing
endowment bezittingen, vermogen
excess demand vraagoverschot, verschil tussen vraag en aanbod (met een minteken als het aanbod groter is dan de vraag)
redistribution herverdeling

The purpose of general equilibrium theory is to understand the process of price formation in exchange economies. Such economies involve a number of "goods" or "assets" which are traded between "agents". Trading is an activity that is as old as mankind, and that is generally beneficial to all parties involved. The conditions of trading are commonly expressed by means of a *price system*; for instance if the price of asset 1 is 10 per unit, and the price of asset 2 is 5 per unit, this means that two units of asset 2 can be traded against one unit of asset 1. Of course the process of trading is usually facilitated by the use of money, but for the purposes of analysis we can think of goods being traded directly against each other. General equilibrium theory attempts to explain prices by setting up equations for equilibrium between supply and demand.

It will be assumed that the number of assets in the economy is finite, say m. A collection of goods is generally called a *consumption bundle*; when the goods are financial assets, it is however more customary to speak of a *portfolio*. A portfolio can then be represented as a vector of *asset holdings* $(x^1, \ldots, x^m)$ (x^j denotes the number of units of asset j in the portfolio). The numbers x^j do not necessarily have to be integers, nor do they necessarily have to be positive.

To formulate the conditions for general equilibrium, we need to know the *demand functions* of agents. These demand functions can be constructed on the basis of *preferences*. We assume that the degree of attractiveness of a particular portfolio to a given agent is measured by a *preference function*. The higher the number associated to a portfolio by the preference function, the more attractive this portfolio is to the agent. Different agents may have different preference functions. A standard assumption is that the preference functions are strictly increasing in all of their arguments ("more is better"), and that they are strictly concave in all of their arguments ("getting something more makes you happier when you have little than when you already have a lot"). In the case of financial assets, concavity may also express a degree of *risk aversion*.

Assume that there are k agents. Each agent has an initial portfolio which is called the agent's *initial endowment*. The quantity of asset j ($j = 1, \ldots, m$) in the initial endowment of agent i ($i = 1, \ldots, k$) is denoted by ω_i^j, so that the initial portfolio of this agent is given by $(\omega_i^1, \ldots, \omega_i^m)$. After trading, the agents in general hold different portfolios $(x_i^1, \ldots, x_i^m)$ ($i = 1, \ldots, k$). A set of portfolios of agents is called an *allocation*. An allocation is said to be *feasible* if the total amount of each good is the same before and after trading:

$$\sum_{i=1}^{k} x_i^j = \sum_{i=1}^{k} \omega_i^j \qquad (j = 1, \ldots, m). \tag{3.1}$$

When the goods are financial assets, an allocation is essentially a *redistribution of risk*. The equations above (one for each asset) are called the *market clearing conditions*.

When the preferences of agents and their initial endowments are specified, we can start the equilibrium analysis. One possible strategy is as follows. First compute the demand function for each agent. When prices of all of the assets are given, the demand function of an agent specifies how much of each good the agent would like to have at the given prices. If an agent's demand for a particular asset is less than the agent's initial endowment of that asset, then effectively a supply is generated. Having computed all demand functions, we can compute the total demand for each asset. Subtracting from this the total initial endowment of this asset, we arrive at the *excess demand function* for each asset. This is still a function of the asset prices that were originally given. The *equilibrium prices* can now be found by solving the equations that arise by requiring that the excess demands for all assets should be zero, with the asset prices as unknowns. The equilibrium that is found in this way is called a *competitive equilibrium*, since it is found by the adjustment of prices in a situation of free trading.

A vector of prices $(\pi_1, \ldots \pi_m)$ for the m assets will be called a *price system*. Since we have assumed that preference assumptions are strictly increasing in their argument so that each asset is indeed a "good" (not a "bad"), only *positive* prices are taken into consideration. Given a price system, the agents determine their demand for assets by solving an *individual optimization problem*. They look for the values x_i^j such that their individual preference function $U_i(x_i^1, \ldots, x_i^m)$ is optimized subject to the budget constraint

$$\pi_1 x_i^1 + \cdots + \pi_m x_i^m \;\leq\; \pi_1 \omega_i^1 + \cdots + \pi_m \omega_i^m. \tag{3.2}$$

The budget constraint (3.2) depends on the chosen prices, so that different prices generate different optimal portfolios. Under the standard assumption that the preference functions are strictly increasing in all of their arguments and given the positivity of prices, the budget constraint (3.2) is always active and so we can take it as an equality constraint:

$$\sum_{j=1}^{m} \pi_j x_i^j = \sum_{j=1}^{m} \pi_j \omega_i^j. \tag{3.3}$$

We then have for agent i ($i = 1, \dots, k$) the optimization problem

$$\begin{aligned} &\text{maximize} \quad && U_i(x_i^1, \dots, x_i^m) \\ &\text{subject to} \quad && \pi_1 x_i^1 + \cdots \pi_m x_i^m = \pi_1 \omega_i^1 + \cdots + \pi_m \omega_i^m. \end{aligned} \tag{3.4}$$

To solve this, we can use the Lagrangian method. Form the Lagrangian function

$$L_i(x_i^1, \dots, x_i^m) = U_i(x_i^1, \dots, x_i^m) + \lambda_i \Big(\sum_{j=1}^{m} \pi_j \omega_i^j - \sum_{j=1}^{m} \pi_j x_i^j \Big).$$

Differentiating, we obtain the m equations

$$\frac{\partial U_i}{\partial x_i^1}(x_i^1, \dots, x_i^m) - \lambda_i \pi_1 = 0, \quad \dots, \quad \frac{\partial U_i}{\partial x_i^m}(x_i^1, \dots, x_i^m) - \lambda_i \pi_m = 0. \tag{3.5}$$

These m equations together with the budget constraint in (3.4) determine the $m+1$ unknowns $x_i^1, \dots, x_i^m, \lambda_i$. Under suitable concavity conditions, there is a unique solution which indeed solves the optimization problem (3.4). The full set of equations for competitive equilibrium is constituted by the budget constraint (3.3), the optimality equations (3.5), and the market clearing condition (3.1). A summary of all ingredients of general equilibrium theory is given in Table 3.1.

The equations of competitive equilibrium as shown in Table 3.1 form a set of $m + k + mk$ equations (partly linear, partly nonlinear) in $m + k + mk$ unknowns. The equations are not independent, however. To see where the dependency comes from, note that by multiplying the market clearing constraint for asset j by π_j and summing over j we obtain the relationship

$$\sum_{j=1}^{m} \sum_{i=1}^{k} \pi_j x_i^j = \sum_{j=1}^{m} \sum_{i=1}^{k} \pi_j \omega_i^j. \tag{3.6}$$

This expresses the fact that, since the amounts of assets do not change by trading, the total value of all assets before trading must be the same as the total value of assets after trading. The value of the portfolio held by agent k can be computed as the total value of all assets minus the total value of the portfolios of agents 1 to $k-1$. Therefore, if the budget constraint is satisfied for agents 1 to $k-1$, then the market clearing condition implies that the budget constraint is also satisfied for agent k. Therefore, given that market clearing is imposed, the equation that expresses the budget constraint for agent k is redundant.

As a consequence of the fact that the equations of general equilibrium are not independent, prices of assets are not determined uniquely. In fact one can see directly from the

Equations of competitive equilibrium — general version

Inputs:

$k =$ number of agents

$m =$ number of assets

$U_i(x^1, \ldots, x^m) \quad (i = 1, \ldots, k)$ preference functions

$\omega_i^j \quad (i = 1, \ldots, k; j = 1, \ldots, m)$ initial endowments

Unknowns:

$\pi_j \quad (j = 1, \ldots, m)$ asset prices

$\lambda_i \quad (i = 1, \cdots, k)$ Lagrange multipliers

$x_i^j \quad (i = 1, \ldots, k; j = 1, \ldots, m)$ equilibrium holdings

Equations:

budget constraint

$$\sum_{j=1}^{m} \pi_j x_i^j = \sum_{j=1}^{m} \pi_j \omega_i^j \qquad (i = 1, \ldots, k)$$

individual optimality subject to budget constraint

$$\frac{\partial U_i}{\partial x_i^j}(x_i^1, \ldots, x_i^m) = \lambda_i \pi_j \qquad (i = 1, \ldots, k; j = 1, \ldots, m)$$

market clearing

$$\sum_{i=1}^{k} x_i^j = \sum_{i=1}^{k} \omega_i^j \qquad (j = 1, \ldots, m)$$

Table 3.1: Equations for competitive equilibrium. The equations determine asset prices only up to a multiplicative constant (choice of unit of currency).

equations that if $(\pi_1, \ldots, \pi_m)$ is a solution, then also $(\alpha\pi_1, \ldots, \alpha\pi_m)$ is a solution where α can be any positive number. This indeterminacy is not surprising. The exchange rate between assets (how many units of one asset can be traded for one unit of another asset) is determined by *relative* prices. Since there is nothing in the equations of general equilibrium which fixes a unit of currency, it should in fact be expected that the solution is only determined up to a proportionality constant. One could arbitrarily choose one of the assets and make the price of that asset equal to 1; this corresponds to the choice of a unit of currency.

The relation (3.6) is known as Walras' law, after the French-Swiss economist who has created the groundwork of general equilibrium theory in the 19th century.[2] Walras in his days was satisfied to note that the number of independent equations is equal to the number of unknowns if one adds one equation to fix the unit of currency. Given that we have a nonlinear system of equations, the mere fact that the number of equations is equal to the number of unknowns does not suffice to guarantee even the existence of a solution. In the mid-20th century, work by mathematical economists including Arrow[3] and Debreu[4] has shown that solutions to the competitive equilibrium equations do exist under very general conditions. In this work it was also shown that the equilibrium solution is Pareto optimal,[5] which means that it is not possible to propose a feasible reallocation that would be preferred by at least one agent and that does not decrease the preference level of any agent (the "first welfare theorem").

Even when one can show theoretically that a solution exists to the equations of general equilibrium, computing the solution in a particular case can be challenging. The following iterative procedure was proposed by Walras:

- a market maker announces prices
- the agents solve their individual optimization problems, and announce their demands for the various assets
- the market maker compares total demand to total supply for each asset, and increases/reduces the prices for assets depending on the difference between demand and supply being positive or negative
- the market maker announces the revised set of prices, and a new iteration is started.

Walras called this procedure *tâtonnement* ("groping around"). The procedure assumes that an unselfish market maker is present, and that no actual trading takes place until the equilibrium has been reached. These assumptions fit in with the idealized nature of the equilibrium

[2] Léon Walras (1834–1910). Walras was born in France but spent most of his academic career at the University of Lausanne in Switzerland.

[3] Kenneth Joseph Arrow (1921), American economist; Nobel prize 1972.

[4] Gérard Debreu (1921–2004), French/US economist; Nobel prize 1983.

[5] Vilfredo Federico Damaso Pareto (1848–1923), Italian economist and philosopher. Pareto was a professor at the University of Lausanne where he worked with Walras.

model itself. In simple examples, it is possible to find the equilibrium by direct computation, without the need for an iterative approximation procedure. Some particular cases are worked out in sections 4.1 and 4.2.

The equations in (3.5) can be written jointly as

$$\begin{bmatrix} \frac{\partial U_i}{\partial x_i^1}(x_i^1, \ldots, x_i^m) \\ \vdots \\ \frac{\partial U_i}{\partial x_i^m}(x_i^1, \ldots, x_i^m) \end{bmatrix} = \lambda_i \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_m \end{bmatrix}. \tag{3.7}$$

The vector that appears on the left hand side here is called the vector of *marginal preferences*. Note that the right hand side does not depend on which agent is involved, except for the scalar multiplication factor λ_i. The above equations can therefore be summarized in words as:

> *at the optimum, the marginal preference vectors of all agents are proportional to the vector of prices.*

This rule can be motivated in words as follows. Agents can change their portfolios by selling one asset and buying another. The proportion in which one asset can be traded against the other is determined by the ratio of the prices of these assets. If the marginal preference vector would not be proportional to the vector of prices, it would be possible to improve the preference level by at least a small amount by a trade that respects the budget constraint.

3.3 Financial assets

English-Dutch vocabulary for Section 3.3			
state of nature	toestand van de natuur	**state of the world**	toestand van de wereld

In the previous section, the traded assets were unspecified; they could be financial assets such as German government bonds or physical goods such as coconuts. In this section, we proceed to a more specific situation which is geared towards financial markets. This will lead to a reformulation of the equations of competitive equilibrium. Our framework of analysis is the single-period model with uncertainty (middle case in Fig. 1.1). It will be assumed that agents are free to trade any contracts they like, in other words, they are not dependent upon a financial market which may offer only a limited range of contracts.

3.3.1 State contracts

As stated in Chapter 1, trading in financial markets is driven by the factor *time* and by the factor *uncertainty*. The factor time does not play a role in the single-period model, but uncertainty does. The presence of uncertainty will be modeled here by assuming that there is an outcome space of n different "states".[6] One of these states will materialize, but agents do not know which one. Agents will receive a state-dependent payoff. The incentive for trading comes from the state dependency of payoffs in combination with a presumed level of risk aversion of agents. If agent A has a high payoff in state 1 but a low payoff in state 2, and for agent B it is the other way around, then it may be a good idea (for *both* of them) to trade, that is, to agree to a contract under which A will pay a certain amount to B if state 1 is realized, and vice versa if state 2 comes up.

A *financial contract* is specified by its payoff vector. If the payoff vectors of the financial contracts that are available for trading in the market are sufficient to span the n-dimensional space, then we speak of a *complete market*. In a complete market, every payoff vector can be realized by a linear combination of payoff vectors associated with available contracts. It is then sufficient, and often convenient, to work with *state contracts* as assets. This means in particular that m, the number of assets, is equal to n, the number of states.

A state contract entitles the holder to a payment of 1 if a given state is realized, and 0 if any other state is realized. A contract between agents A and B as described in the previous paragraph can then be viewed as a trade in which agent A delivers a certain number of units of the state contract for state 1 to B in exchange for a certain number of units of the state contract of state 2. Initial endowments which deliver a state-dependent payoff can be viewed as portfolios of state contracts. For instance, in the situation of Example 3.1.1 one can define the asset S as the state contract that pays 1 on a sunny day and 0 on a rainy day, and one can introduce the asset R as the state contract that pays 1 on a rainy day and nothing on a sunny day. The initial endowment of Alice can then be described as the portfolio consisting of 20 units of state contract S and 5 units of state contract R.

3.3.2 Expected utility

The setting in which assets are actually payoffs that occur in different states makes it possible to construct preference functions in a particular way. Let $u(x)$ be a *utility function*, that is, an increasing and concave function of one variable. Also, assume that for each state j a positive number p_j is given, with $\sum_{j=1}^{n} p_j = 1$. One can then define a preference function $U(x^1, \ldots, x^n)$ as follows:

$$U(x^1, \ldots, x^n) = \sum_{j=1}^{n} p_j u(x^j). \tag{3.8}$$

[6] The terms "state of nature" or "state of the world" are sometimes used instead of just "state".

So, instead of having to specify a function of n variables, we now only need to specify a function of one variable and a collection of n positive constants.

Several interpretations of the numbers p_j are possible. The most straightforward interpretation is to look at p_j as the *objective probability* of the occurrence of state j.[7] The expression on the right hand side of (3.8) is then the *expected utility* associated to the portfolio $(x^1, \dots, x^n)$. For this reason, the framework in which preference functions are described as in (3.8) is called the *expected utility framework*. However, given the fact that in economics we usually do not have the luxury of being able to repeat experiments many times under identical circumstances, the notion of "objective" probability may not be entirely unshakable. Instead one may think of the p_j's as representing *subjective probabilities*, which just express the beliefs of one particular agent. The right hand side of (3.8) is then still an expected utility, but the expectation is now subjective. Thirdly the numbers p_j may be viewed as just *weights* which are attached to the states by a given agent. Agents may deliberately choose weights that are different from their subjective beliefs, for instance as a way of expressing aversion to particular states. The requirement that the weights should add to 1 can then be viewed as just an arbitrary normalization. However, the expected utility interpretation either in the objective or in the subjective sense is more common.

Consider an agent whose preferences across portfolios of state contracts is given by the expression above, where $u(x)$ is a continuously differentiable, strictly increasing, and strictly concave function. The partial derivative of $U(x^1, \dots, x^n)$ with respect to x^j (payoff in state j) is given by

$$\frac{\partial U}{\partial x^j}(x^1, \dots, x^n) = p_j u'(x^j).$$

Therefore, the optimality conditions in (3.5) can be written in the form

$$p_j u'(x^j) = \lambda \pi_j \qquad (j = 1, \dots, n). \tag{3.9}$$

Strict concavity of the function $u(x)$ means that its derivative $u'(x)$ is strictly decreasing. Therefore the derivative has an inverse function that will be denoted by $v(x)$; in other words, the function $v(x)$ is defined by

$$v(u'(x)) = x \qquad \text{for all } x.$$

The derivative of the utility function, $u'(x)$, is often called the *marginal utility*, and correspondingly the function $v(x)$ is called the *inverse marginal utility*. In terms of the inverse marginal utility, the condition (3.9) can be rewritten as

$$x^j = v(\lambda \pi_j / p_j). \tag{3.10}$$

[7] Here we encounter a notational conflict between probability theory and microeconomics. In microeconomics, the letter p is used for prices, whereas in probability theory it stands for probability. The finance literature largely follows the notational conventions of probability theory rather than the conventions of microeconomics. In these notes we use p for probabilities, and we will use π for prices.

The Lagrange multiplier λ that appears here can be solved from the budget constraint (3.2) which, by virtue of (3.10), can be written as

$$\sum_{j=1}^{n} \pi_j v(\lambda \pi_j / p_j) = \sum_{j=1}^{n} \pi_j \omega^j. \tag{3.11}$$

Therefore, one way to obtain an explicit expression for the demands x^j in terms of the price system $(\pi_1, \ldots, \pi_n)$ is to solve λ from (3.11) and to insert the result into (3.10).

Some popular utility functions are given in Table 2.1. All of these utility functions satisfy the conditions that were imposed above (continuously differentiable, strictly increasing, strictly concave). The parameter γ that appears in the power utility function is a *risk aversion parameter*: higher values of γ correspond to stronger risk aversion. On the other hand the parameter τ that appears in the exponential utility function is a *risk tolerance parameter*; higher values of τ correspond to *less* risk aversion. There is some freedom in choosing a parametrization, since utility functions can be shifted up or down and scaled without any effect on the optimal decisions. It is customary to choose the parametrization in such a way that the expression for marginal utility becomes as simple as possible.

3.3.3 Risk-adjusted probabilities

Suppose that we do have objective probabilities p_j $(j = 1, \ldots, n)$ associated to the states in our single-period model. The expected payoff of the state contract that goes with state j is then equal to p_j units of currency. Therefore, the number p_j might be thought of as a reasonable candidate for the price of a state contract. However, it turns out that the state prices obtained from general equilibrium theory are not necessarily proportional to the state probabilities. The relation between state probabilities and state prices is in fact one of the main topics of investigation in financial theory.

As noted before, the equations of general equilibrium theory only determine prices up to a proportionality factor, because of the lack of a prescribed unit of currency. In the single-period model that we are discussing in this section, there *is* in fact a unit of currency, namely the one that is used for the payoffs. Still the equations of general equilibrium only determine prices up to a proportionality constant. The reason is that in principle the currency used to express prices of contracts can be different from the currency in which payoffs are expressed. Even when the currencies are the same, the application of a *discount factor* may be reasonable when contracts are paid for at a different moment in time than when payoffs are received. We can however introduce normalized state prices q_j which by definition add up to 1 in the following way:

$$q_j = \frac{\pi_j}{\sum_{j=1}^{n} \pi_j}.$$

These normalized state prices are like probabilities in the sense that they are positive and their sum is 1. Since the numbers q_j are influenced by the risk aversion of agents in the economy, they are often referred to as *risk-adjusted probabilities*. It should be noted that the risk-adjusted probabilities may be determined even when objective probabilities are not given, since the equations of general equilibrium can also be solved when only subjective probabilities are specified. Therefore, risk-adjusted probabilities can exist even when objective probabilities do not exist.

The numbers q_j are also called *risk-neutral* or *implied* probabilities. The term "risk-neutral" comes from the fact that, if the competitive equilibrium would be solved with objective probabilities and risk-neutral agents (utility function $u(x) = x$), then the q_j's would coincide with the p_j's. The reason that the term *implied* probability is used is that, if the preference functions of agents are unknown but market prices of portfolios can be observed, it is possible to extract state prices from market data rather than from an equilibrium computation. This methodology will be illustrated in Chapter 7.

It is often convenient to use notation and terminology in which the q_j's are treated as actual probabilities. For instance, the price of an asset that pays c^j in state j is given by $\sum_{j=1}^n c^j \pi_j$. We can write this as the pricing formula

$$\pi(C) = \alpha E^Q[C]$$

where $C = (c^1, \dots, c^n)$, $\pi(C)$ is the price of C, and the symbol "E^Q" indicates that expectation is taken with the q_j's as probabilities of the states. The multiplicative constant α can be be written as $\pi(\mathbb{1})$ where $\mathbb{1}$ denotes the portfolio $(1, \dots, 1)$; in other words, $\mathbb{1}$ is the asset that pays 1 irrespective of the state that is realized. In this way we arrive at the pricing formula

$$\pi(C) = \pi(\mathbb{1}) E^Q[C]. \tag{3.12}$$

This formula by itself expresses no more than the definition of the probability measure Q and the linearity rule of pricing (the price of a portfolio can be computed as the sum of the prices of the individual assets per unit, times their portfolio weights). It is just a convenient way of writing prices in terms of Q and a multiplicative constant. The key issue is to determine Q.

The equations of competitive equilibrium in the particular case that was considered in this section are summarized in Table 3.2. Because of the normalization constraint that has been added, the number of equations is now equal to the number of unknowns, and under suitable circumstances a unique solution exists. Some examples are shown in Chapter 4.

3.4 Good states and bad states

English-Dutch vocabulary for Section 3.4	
comonotonic comonotoon, in gelijke richting gaand	**comonotonicity** comonotoniciteit

Equations of competitive equilibrium — version with utility functions, state contracts, objective probabilities, and state prices expressed in terms of risk-adjusted probabilities

Inputs:

$k =$ number of agents
$n =$ number of states
$u_i(x) \quad (i = 1, \ldots, k)$ utility functions
$p_j \quad (j = 1, \ldots, n)$ objective probabilities
$\omega_i^j \quad (i = 1, \ldots, k; j = 1, \ldots, n)$ initial endowments

Unknowns:

$q_j \quad (j = 1, \ldots, n)$ risk-adjusted probabilities
$\lambda_i \quad (i = 1, \cdots, k)$ Lagrange multipliers
$x_i^j \quad (i = 1, \ldots, k; j = 1, \ldots, n)$ equilibrium payoffs

Equations:

budget constraint

$$\sum_{j=1}^{n} q_j x_i^j = \sum_{j=1}^{n} q_j \omega_i^j \qquad (i = 1, \ldots, k)$$

individual optimality subject to budget constraint

$$u_i'(x_i^j) = \lambda_i \frac{q_j}{p_j} \qquad (i = 1, \ldots, k; j = 1, \ldots, n)$$

market clearing

$$\sum_{i=1}^{k} x_i^j = \sum_{i=1}^{k} \omega_i^j \qquad (j = 1, \ldots, n)$$

normalization

$$\sum_{j=1}^{n} q_j = 1$$

Table 3.2: Equations of competitive equilibrium. The case in which the agents use individual (subjective) probabilities of the states can be obtained from the above by replacing p_j by p_j^i.

In a situation in which the preferences of agents are given by utility functions and all agents use the same (objective) probabilities, the relation (3.9) shows that the quotient of the risk-adjusted probability q_j and the objective probability p_j is, up to a proportionality factor, equal to the marginal utility of agents in their respective payoffs when state j occurs. Indeed, the equations (3.9) must in equilibrium be satisfied for all individual agents. The solvability of the equilibrium conditions guarantees that this is indeed possible; of course, it requires suitable choice of the prices as well as of the payoffs received by the agents. The optimality condition (3.9) that must be satisfied for each agent i can be rewritten as

$$\frac{q_j}{p_j} \propto u_i'(x_i^j) \qquad (i = 1, \ldots, k) \tag{3.13}$$

The proportionality symbol $\propto$ indicates that the ratio of the left hand side and the right hand side is a constant independent of the index j which refers to the event j; the proportionality constant may however depend on the index i which is used to indicate the agent.

The left hand side of (3.13) does not depend on the index i. From that it follows that, in equilibrium, the payoffs of the agents are ordered in same way; that is, if one agent gets a higher payoff in state 1 than in state 2, then state 1 is in fact better than state 2 for *all* agents. This holds by the standard assumption that the marginal utility functions $u_i'(x)$ of all agents are decreasing. Therefore, in equilibrium, the agents agree on the ordering of the states from “good” to “bad”. This may be expressed by saying that the equilibrium payoffs are *comonotonic*. In case the agents use different (subjective) probabilities, then the comonotonicity property is not guaranteed.

If the payoffs of all agents are better in state j than state ℓ, then the sum of all payoffs must be higher in state j than in state ℓ; in other words, the ordering of the states is determined by the sums of the payoffs of the agents in each of the states. To state this argument more formally, note that (3.13) validates the following implication, for each agent i:[8]

$$\frac{q_j}{p_j} > \frac{q_\ell}{p_\ell} \Rightarrow u_i'(x_i^j) > u_i'(x_i^\ell) \Rightarrow x_i^j < x_i^\ell.$$

Since this holds for all agents, we can therefore also say

$$\frac{q_j}{p_j} > \frac{q_\ell}{p_\ell} \Rightarrow \sum_{i=1}^k x_i^j < \sum_{i=1}^k x_i^\ell.$$

In equilibrium, market clearing must hold, so that $\sum_{i=1}^k x_i^j = \sum_{i=1}^k \omega_i^j$ and likewise for ℓ. Therefore,

$$\frac{q_j}{p_j} > \frac{q_\ell}{p_\ell} \Rightarrow \sum_{i=1}^k \omega_i^j < \sum_{i=1}^k \omega_i^\ell. \tag{3.14}$$

Note that the two sums on the right hand side denote the total income of all agents together in state j and in state ℓ respectively. Therefore, in equilibrium, the ratios q_j/p_j are ordered

[8] It is assumed that the utility functions of all agents are strictly increasing and strictly concave, so that the marginal utility functions are positive and strictly decreasing.

in a way that is opposite to the ordering of total income (also referred to as *aggregate income*). In particular, this means that the implication (3.14) can also be used in the opposite direction. If we look at the general equilibrium solution as a way of dividing aggregate income (which in general is different in different states) among a group of agents, then the relation (3.14) states that the reward per unit of contributed income is higher in states where income is scarce than in states where income is plentiful.

In the special situation in which the aggregate income is the same across events, it follows from the implication (3.14) that $q_j/p_j = q_\ell/p_\ell$ for all j and ℓ in $\{1,\dots,m\}$. Because both the q_j's and the p_j's must sum to 1, it follows that we must have $q_j = p_j$ for all j. Moreover, it follows that in this case $u_i'(x_i^j) = u_i'(x_i^\ell)$ for all j and ℓ in $\{1,\dots,m\}$; since the utility functions are assumed to be strictly increasing, this implies that $x_j^i = x_\ell^i$ for all j and ℓ. In other words, in equilibrium each agent i gets the same payoff in every state (but these payoffs may differ between agents). In summary, this might be phrased as: *in an economy populated by risk averse expected utility agents who agree on the probabilities of the states, absence of risk in aggregate income implies that in equilibrium no individual bears risk.*[9]

Let us return now to the situation in which aggregate income does vary across states. Because the marginal utilities are *decreasing*, q_j is *smaller* than p_j when j is a *good* state (in the sense that aggregate income is high), and q_j is *larger* than p_j when j is a *bad* state. This allows us to say more about the nature of the adjustment in the adjusted probabilities q_j: the adjustment is *upward* for states that are *bad* (in equilibrium, i.e. after risk sharing), and the adjustment is *downward* for *good* states. So the risk-adjusted probabilities of bad states are higher than their actual probabilities; one might say that the q_j's represent a *pessimistic* view. This result is a consequence of the assumption that marginal utilities are decreasing, which is an expression of *risk aversion* (concave utilities).

The ratio q_j/p_j is a function of the state j. The relation (3.14), however, shows that if aggregate income in two states j_1 and j_2 is equal, then the corresponding ratios q_{j_1}/p_{j_1} and q_{j_2}/p_{j_2} must also be equal. In other words, the ratio q_j/p_j can also be viewed as a function of aggregate income. More specifically, as discussed above, it is a *decreasing* function of aggregate income.

3.5 Testable implications

The model of financial markets that has been developed above represents an attempt to explain price formation in financial markets as a process of risk sharing. The key elements are as follows:

- the "goods" (traded assets) are payoffs in different states of the world

[9] Similar reasoning can be applied to the situation in which we have two states in which aggregate income is the same; see Exc. 6.

English-Dutch vocabulary for Section 3.5

excess return overrendement	**net return** nettorendement
fee vergoeding, betaling voor diensten	**return** rendement
gross return brutorendement	

- prices are determined by supply and demand
- supply and demand, in their turn, are driven by economic activity ("initial endowments") and by agents' risk aversion.

The model includes all of these elements; on the other hand, it is a rather simplified model compared to the complexity of actual financial markets. The question arises whether the model can be put to the test, in the sense that predictions that follow from it can be either verified or refuted on the basis of empirical data that are generated by the markets.

It is not a straightforward exercise to confront the model with the real world. One problem for instance is that the finite set of possible future states, which is assumed in the theory, is not actually available. In order to arrive at statements that have a recognizable real-world form, it is useful to rewrite the formulas that have been derived above. The pricing formula (3.12) can be written as[10]

$$\pi(C) = E[MC] \tag{3.15}$$

if we define

$$M_j = \pi(\mathbb{1})\frac{q_j}{p_j}. \tag{3.16}$$

The random variable M that is defined in this way is called the *pricing kernel*. It is the product of two factors, of which one corresponds to the effect of time (this is $\pi(\mathbb{1})$, the *discount factor*), and the other relates to the effect of uncertainty (the quotient q_j/p_j). It is convenient to introduce the *normalized pricing kernel*, which includes only the second factor. This state-dependent quantity is denoted by θ:

$$\theta_j = \frac{q_j}{p_j}\,. \tag{3.17}$$

Such a quantity, which is obtained as quotient between two sets of probabilities so that $E^Q[Z] = E[\theta Z]$ for all Z, is also called a *Radon-Nikodym derivative*. Since by definition the q_j's sum to 1, the relation

$$E[\theta] = 1 \tag{3.18}$$

must hold. Under the assumption of "objective probabilities", i.e. all agents use the same probabilities to compute their expected utilities, it has been derived in the previous section that the normalized pricing kernel is a decreasing function of aggregate income.

[10]The symbol "E" is used for expectation with respect to the real-world probabilities; in other words, for any random variable Z defined on $\Omega = \{1, \dots, n\}$, we have $E[Z] = \sum_{i=1}^n p_j z_j$. Expectation with respect to the risk-adjusted probabilities is denoted by E^Q, so that $E^Q[Z] = \sum_{i=1}^n q_j z_j$.

For a given asset with payoff C at time 1, the random variable $C/\pi(C)$ is called the *return* of the asset.[11] It is the amount of payoff generated by the asset, relative to the size of the investment made at time 0. As viewed from time 0, the return is a random quantity. The pricing formula (3.12) can be rewritten as a return formula by dividing both sides of the equation by $\pi(C)$:

$$1 = E\Big[M\frac{C}{\pi(C)}\Big]. \tag{3.19}$$

From an econometric point of view, to make data from different periods more comparable, it is better to work with differences of returns rather than with absolute returns. For instance one can take the difference with the return on the riskless asset. The return on this asset will be denoted by R^f; by definition, we have

$$R^f = \frac{1}{\pi(\mathbb{1})}. \tag{3.20}$$

The riskless return R^f is also written as $1 + r$, where r is the simply compounded interest rate for the period between time 0 and time 1. By definition of the normalized pricing kernel, the relation

$$E[\theta R] = R^f \tag{3.21}$$

holds for any (in general risky) return $R = C/\pi(C)$. The difference between a given (risky) return R and the riskless return R^f is called the *excess return* and is often denoted by R^e. The excess return is a random variable. From (3.19), we can write

$$E[MR^e] = 0 \tag{3.22}$$

which is equivalent to

$$E[\theta R^e] = 0. \tag{3.23}$$

Given the standard formula for the covariance of two random variables, and given also the relation $E[\theta] = 1$, another equivalent statement is

$$E[R^e] = -\operatorname{cov}(\theta, R^e). \tag{3.24}$$

The quantity at the left hand side is the *expected excess return*, which can be determined from time series data on asset prices (under a stationarity assumption). The right hand side cannot be directly obtained from experimental data, though, since the quantity θ is not observable as such.

However, the theory also states that (at least when agents do not disagree too much about the probabilities of future events) the normalized pricing kernel θ is related counter-monotonically to aggregate income; in other words (see (3.14)), "good" states (i.e. states

[11] This is also sometimes called the *gross return*, and then the *net return* is defined as the gross return minus 1. The terms "gross return" and "net return" are used in practice however also to distinguish between returns before and after subtraction of various costs such as taxes and fees; so care should be taken not to confuse terms.

with high aggregate income) are associated with low realizations of θ, and "bad states" are associated with high realizations of θ. Aggregate income is also referred to as the *market portfolio*. The return on the market portfolio is denoted by R^m. This quantity may be thought of as being observable; in empirical work, it is often taken to be given by the return on a broad stock index such as the Dow Jones index.[12]

According to the theory, the normalized pricing kernel θ and the return on the market portfolio R^m should be negatively correlated. As a first-order approximation, a linear relationship may be suggested:

$$\theta = a - bR^m \tag{3.25}$$

where a and b are both positive. The coefficients a and b must be chosen in such a way that $E[\theta] = 1$; this means that $a - bE[R^m] = 1$, or in other words $a = 1 + bE[R^m]$. The relationship above can therefore be written as

$$\theta = 1 - b(R^m - E[R^m]). \tag{3.26}$$

Under the assumption (3.25), the relation (3.24) that is predicted by the theory can be written in more real-life terms as

$$E[R^e] = b \operatorname{cov}(R^e, R^{me}). \tag{3.27}$$

This is a relationship that should hold for every excess return, with a coefficient b that does not depend on the asset that generates the return. In particular the relation should also hold for the market portfolio:

$$E[R^{me}] = b \operatorname{var}(R^{me}). \tag{3.28}$$

From this we find that $b = E[R^{me}]/\operatorname{var}(R^{me})$, so that the equation (3.27) can now be written more specifically as

$$E[R^e] = \frac{\operatorname{cov}(R^e, R^{me})}{\operatorname{var}(R^{me})} E[R^{me}]. \tag{3.29}$$

This is the well known Capital Asset Pricing Model, that has found extensive use as a way of expressing what kind of return should be expected when a certain risk is taken. The model can be rewritten in the form

$$E[R] = R^f + \beta\big(E[R^m] - R^f\big) \tag{3.30}$$

where

$$\beta = \frac{\operatorname{cov}(R^e, R^{me})}{\operatorname{var}(R^{me})} \tag{3.31}$$

is known as the *CAPM beta*. To the extent that a suitable proxy for the market portfolio can be found, the relation (3.29) is in a form that can be empirically tested.

[12] It may be debated however to what extent an equity exchange index can be really representative for the notion of "aggregate income" as it is used in the theory, since this notion should also include for instance income from labor, returns from bonds and non-listed equity, harvests from forests and farmlands, and so on.

The CAPM states that the expected excess return of any asset is proportional to its covariance with the market portfolio.[13] This may be compared to the equation (3.24) derived from the theory, which states that the expected excess return of any asset is equal to minus the covariance of its return with the normalized pricing kernel. The CAPM claim is more practical, but it does not follow from the theory without further assumptions. It is conceivable that the normalized pricing kernel can be approximated more closely by a linear combination of the returns of *several* assets. This leads to representations of the form

$$\theta = 1 - \sum_{k=1}^{K} c_k (R_k^e - E[R_k^e]). \tag{3.32}$$

From the above equation and (3.24) we get the *multifactor model*

$$E[R^e] = \sum_{k=1}^{K} c_k \operatorname{cov}(R^e, R_k^e) \tag{3.33}$$

where the c_k's are to be determined.

3.6 Exercises

1. Consider an economy in which there are only two assets. We shall refer to the assets as "cookies" and "lemonade". In this economy there is an agent named Alice. Alice's preference function is given by

$$U_A(x^{\mathrm{C}}, x^{\mathrm{L}}) = \sqrt{x^{\mathrm{C}}} + 2\sqrt{x^{\mathrm{L}}}$$

where x^{C} is the number of cookies she has, and x^{L} refers the number of glasses of lemonade that she owns.[14]

a. In a cookies/lemonade diagram, draw some of Alice's indifference curves. (Indifference curves may also be called equi-preference curves; they connect points which have the same preference value.)

b. If Alice has four cookies and two glasses of lemonade, how many cookies would she marginally trade for one glass of lemonade? ("Marginal" means that she is trading a small amount of cookies against a small amount of lemonade.)

c. Compare Alice's marginal exchange rate of cookies against lemonade at different points on the same indifference curve. When she has more lemonade, is the amount of cookies she

[13] This statement implies that the proportionality constant does not depend on the asset in question.

[14] The terms "cookies" and "lemonade" are used here for ease of interpretation. Alternatively one may think of a situation in which one of two states will materialize, and C and L are the associated state contracts. Alice's preference function can then be interpreted as reflecting expected utility, with utility function $u(x) = \sqrt{x}$, and probabilities $\frac{1}{3}$ and $\frac{2}{3}$ for state C and state L respectively.

wants to trade for even more lemonade higher or lower? Does the answer agree with what you would consider reasonable behavior?

d. Suppose that Alice has a given number of cookies and a given number of glasses of lemonade, and that she can trade cookies for lemonade or vice versa at a given fixed exchange rate. Give a graphical interpretation in the cookies/lemonade diagram of her optimal choice.

2. In the situation of Exc. 1, assume that there is a second agent named Bob. Bob's preference function is

$$U_B(x^{\mathrm{C}}, x^{\mathrm{L}}) = \sqrt{x^{\mathrm{C}}} + \sqrt{x^{\mathrm{L}}}$$

where x^{C} and x^{L} refer to the numbers of cookies and glasses of lemonade held by Bob. Now suppose that Alice has 20 cookies and Bob has 5 glasses of lemonade. They can then consider to exchange some of their assets. This means they will enter into a contract in which Alice hands a certain number of cookies to Bob, and Bob transfers a certain amount of lemonade to Alice. We assume that the market clearing condition is obeyed, so that the total amounts of cookies and of lemonade held by Alice and Bob together are the same before and after the contract.

a. Given the initial endowments as just specified, all possible contracts that clear the market can be indicated by points in the cookies/lemonade diagram for Alice. In this diagram, indicate the region of all contracts that are as good as or better for Alice than her initial situation. Also indicate this region for Bob.[15] Do the two regions overlap?

b. Find the equilibrium prices of cookies and lemonade.

c. In the cookies/lemonade diagram, identify the point that corresponds to the equilibrium contract. Draw the indifference curves for Alice and for Bob that pass through this point, and verify that Pareto optimality holds (i.e. the equilibrium point is the best choice for Alice in the region of pints that for Bob are just as good as or better than the equilibrium, and vice versa). Also draw the line through this point that represents all contracts that correspond to the equilibrium prices.

3. In the situation of Exc. 2, suppose that the dog eats one of Alice's cookies before Alice and Bob have had the chance to trade. How does this affect the relative prices of cookies and lemonade?

4. Consider again the situation of Exc. 2, but now suppose that Alice and Bob do not trade with each other directly. Instead, there is an intermediary named Carol, who initially has no cookies and no lemonade. When Carol announces a certain price of lemonade in terms

[15] Note that, under the assumption that no cookies are lost and no lemonade is spilled, any assignment of cookies and lemonade to Alice also implies an assignment of cookies and lemonade to Bob. The resulting diagram in which the holdings of both Alice and Bob are both represented is called an *Edgeworth box*, after the Irish philosopher and economist Francis Ysidro Edgeworth (1845–1926).

of cookies, both Alice and Bob will trade part of their assets with Carol so as to optimize their preference functions. Carol does not need to offer the same price of lemonade to Alice and to Bob; she can use a "bid price" and an "ask price." The amount of cookies that Alice sells to Carol must be at least be as large as the amount of cookies that Carol sells to Bob, but equality does not need to hold; if there is a surplus, then Carol simply keeps this. An analogous statement holds for lemonade. Suppose that Carol doesn't care for lemonade but that she does like cookies. How should she choose the bid price and the ask price for lemonade in such a way that she maximizes the amount of cookies that she can keep after the trades have been effectuated?

5. In the situation of Exc. 4, suppose that Carol has even stronger negotiation power. She can offer deals to both Alice and Bob, and they can only choose between accepting or not accepting the deal. Assume that Alice and Bob will accept any deal that improves their preference function. How many cookies can Carol now appropriate for herself?

6. Consider a single-period general equilibrium model in which the preferences of all agents are of the expected utility type. The utility functions of different agents need not be the same, but the agents do agree on the probabilities. All utility functions are supposed to be differentiable and strictly concave. Prove that, when aggregate income in two events is the same, the equilibrium allocations in these two events are also the same. (In other words, in equilibrium no individual bears risk across states that are the same in terms of aggregate income. Or, even more briefly: risk is not borne unnecessarily. The statement can also be used to argue that, for the purposes of single-period competitive equilibrium theory under objective probabilities, two events that produce the same aggregate income can effectively be considered as a single event.)

7. Here is an example of economic modeling, based on a recent paper by Van Binsbergen *et al.* that was constructed to give insight in fair risk sharing between younger and older generations within a collective pension fund. The model can be described in words as follows.

Two moments in time are considered: current time (time 0) and a future time (time 1). There are two states that may be realized at time 1: a "good" state H and a "bad" state L. In the pension fund there are two generations, called "young" (Y) and "old" (O). Two assets are available for investment: a risky asset R and a non-risky asset N. A unit of the risky asset pays off R_H units of currency in state H, and R_L units of currency in state L, where $R_H > R_L$. The risky asset is traded on the world market in which there is plenty of liquidity, and therefore its price is determined in this market; the generations in the pension fund do not have any effect on it. By normalization, we can set the price of one unit of the risky asset equal to 1 unit of currency.

At time 0, the members of the young generation have wealth W_Y (amount of money available for investment; assumed to be the same for all members of the young generation). The members of the old generation have wealth W_O, also assumed to be the same for all members of this generation. In addition to the payoffs of their investments, the members of the young generation receive y units of currency (their salary, again the same for everyone) at time 1, irrespective of whether the "good" or the "bad" state is realized. The members of the old generation are retired and do not receive a salary. All agents choose their portfolios in order to optimize the expected utility of payoffs received at time 1. They all use the same probabilities for the "good" and the "bad" state (p_H and p_L respectively, with $p_H + p_L = 1$) and they all have the same utility function $u(\cdot)$.

In terms of equations, the model is described as shown below. All quantities mentioned in the model description, except for the price of the non-risky asset, are considered to be given.

a. Explain the following equations:

$$\pi_L R_L + \pi_H R_H = 1 \tag{3.34}$$

$$\pi_L + \pi_H = \pi_N \tag{3.35}$$

where π_N denotes the (yet to be determined) equilibrium price of one unit of the riskless asset. What is the meaning of the symbols π_L and π_R?

b. The variables x_Y^L, x_Y^H, x_O^L, x_O^H are defined by

$$x_Y^L = h_Y^R R_L + h_Y^N + y \tag{3.36}$$

$$x_Y^H = h_Y^R R_H + h_Y^N + y \tag{3.37}$$

$$x_O^L = h_O^R R_L + h_O^N \tag{3.38}$$

$$x_O^H = h_O^R R_H + h_O^N \tag{3.39}$$

where h_Y^R and h_Y^N represent the portfolio holdings of risky and non-risky assets respectively for members of the young generation, and h_O^R and h_O^N represent the same for members of the old generation. What is the meaning of the variables x_Y^L, x_Y^H, x_O^L, and x_O^H?

c. Explain the following conditions:

$$\begin{aligned} &p_L u(x_Y^L) + p_H u(x_Y^H) \to \text{max} \\ &\qquad\qquad \text{subject to } \pi_L x_Y^L + \pi_H x_Y^H = W_Y + \pi_N y \end{aligned} \tag{3.40}$$

$$\begin{aligned} &p_L u(x_O^L) + p_H u(x_O^H) \to \text{max} \\ &\qquad\qquad \text{subject to } \pi_L x_O^L + \pi_H x_O^H = W_O. \end{aligned} \tag{3.41}$$

d. Explain the following equation:

$$\alpha h_Y^N + (1 - \alpha) h_O^N = 0 \tag{3.42}$$

where α denotes the fraction of young participants in the fund (a given quantity). Why is there no similar equation for the risky asset?

e. The conditions (3.40) and (3.41) are given as optimization problems. Rewrite these conditions as equations, under the standard assumption that the utility function is strictly increasing, strictly concave, and twice continuously differentiable.

f. Count the number of equations that now have been stated, as well as the number of unknowns. Do you expect that there is a unique solution?

8. **a**. Let X be a normal random variable with expectation μ and variance σ^2, and let g be a differentiable function such that $E[g'(X)]$ is finite. Use integration by parts to show that[16]

$$E[(X-\mu)g(X)] = \sigma^2 E[g'(X)]. \tag{3.43}$$

b. Suppose that $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are jointly normally distributed variables with correlation coefficient ρ. Prove that

$$E[(X_1-\mu_1)g(X_2)] = \operatorname{cov}(X_1, X_2)E[g'(X_2)]. \tag{3.44}$$

[Hint: write $X_1 - \mu_1 = \rho\frac{\sigma_1}{\sigma_2}(X_2 - \mu_2) + X_3$ where X_3 and X_2 are independent, and $EX_3 = 0$. Then use part a.]

c. In the same situation as in the previous question, show that

$$\operatorname{cov}(X_1, g(X_2)) = \operatorname{cov}(X_1, X_2)E[g'(X_2)]. \tag{3.45}$$

d. Now consider an economy in which the market return R^m follows a normal distribution[17] with expectation μ_m and variance σ_m^2. Suppose that the return R of a given asset is also normally distributed, with expectation μ and variance σ^2, and that R and R^m are jointly normally distributed with correlation coefficient ρ. Under the assumption that the normalized pricing kernel can be written as a function of the market return, show that the CAPM beta is exact for R; in other words, show that

$$\frac{E[R^e]}{E[R^{me}]} = \frac{\operatorname{cov}(R, R^m)}{\operatorname{var}(R^m)}\,.$$

[Note that the assumption (3.25) is not used; on the other hand, rather strong distributional assumptions are made.]

[16]The result in (3.43) is called *Stein's lemma*, after the American statistician Charles Stein (1920).

[17]Obviously this is not a finite-state economy as discussed in the chapter; however, one can still use the main formulas, such as the covariance expression (3.24).

Chapter 4

Equilibrium: special cases

4.1 Log utility

In this section we consider a market for single-period financial assets in which all investors have logarithmic utility. The investors may be different though in their assessments of the probabilities of the various possible states. Under the usual assumption that the assets that we consider are state contracts, the preference function of agent i is then given by

$$U(x_i^1, \ldots, x_i^n) = \sum_{j=1}^{n} p_j^i \ln x_i^j$$

where the numbers p_i^j may depend on i. These numbers can therefore be viewed as *subjective probabilities*: p_j^i is what agent i believes that the probability of state j is. We will make the assumption that $p_j^i > 0$ for all i and j, so that it doesn't happen that there are some states which some agents take to be impossible whereas other agents do think of them as possible. This can be classified as an efficiency assumption.

To solve the model, we can proceed as follows. The optimality equations for agent i are (cf. (3.9))

$$p_j^i \frac{1}{x^j} = \lambda_i \pi_j \qquad (j = 1, \ldots, n). \tag{4.1}$$

The Lagrange multiplier λ_i can be obtained from the budget constraint:

$$\sum_{j=1}^{n} \pi_j v(\lambda_i \pi_j / p_j^i) = \sum_{\ell=1}^{n} \pi_\ell \omega_i^\ell. \tag{4.2}$$

where in this case $v(x) = 1/x$ (see Table 2.1). In the right hand side, the usual summation index j is replaced by ℓ in order to avoid confusion with the index j that is used later on. Because the inverse marginal utility is given by $v(x) = 1/x$, the price π_j drops out of the left hand side of this equation; making use also of the fact that the probabilities p_j^i must sum to 1, we conclude from (4.2) that

$$\lambda_i = \Big(\sum_{\ell=1}^{n} \pi_\ell \omega_i^\ell \Big)^{-1}.$$

Inserting this into (4.1), we obtain the demand function of agent i for asset j:

$$x_i^j(\pi_1, \ldots, \pi_n) = \frac{p_j^i}{\pi_j} \sum_{\ell=1}^n \pi_\ell \omega_i^\ell. \tag{4.3}$$

This formula indicates that the demand of agent i for asset j will be higher when the agent believes that the outcome j is more likely (recall that assets are assumed to be state contracts, so that asset j is the contract that pays one unit of currency when outcome j appears), when the price of asset j is lower, and when the value of the agent's initial endowment is higher (in other words when the agent has more to spend). All of this appears rather reasonable. The formula also shows that the demand of agent i for asset j depends not only on the price of asset j itself, but also on the prices of all other assets since these have an effect on the value of the agent's initial endowment. In terms of the adjusted probabilities, (4.3) can be rewritten as

$$x_i^j = \frac{p_j^i}{q_j} \sum_{\ell=1}^n q_\ell \omega_i^\ell = \frac{p_j^i}{q_j} E^Q \omega_i. \tag{4.4}$$

This expresses the equilibrium payoff of agent i in state j as the average payoff (weighted by the *adjusted* probabilities) of the agent's initial endowment, times a factor which is equal to the ratio between the subjective probability of state j according to agent i and the adjusted probability of that state. The multiplication factors corresponding to the various states are on average equal to 1, if we computed a weighted average with the adjusted probabilities as weights:

$$\sum_{j=1}^n q_j \frac{p_j^i}{q_j} = \sum_{j=1}^n p_j^i = 1.$$

From this we see that $E^Q x_i = E^Q \omega_i$, which verifies that the budget constraint is indeed satisfied.

In the case of *objective* probabilities, the equation (4.4) shows that the equilibrium portfolios of the agents are all multiples of each other. Because the sum of the portfolios of the agents must be equal to the market portfolio which pays $\sum_{i=1}^k \omega_i^j$ in state j, it follows that the equilibrium portfolios of the agents are fixed fractions of the market portfolio. Therefore we can already say that, under log utility and objective probabilities, the equilibrium allocation assigns to each agent a fixed percentage of aggregate income. In other words, we obtain a *proportional* allocation rule. The competitive equilibrium for log utility agents under objective probabilities is summarized in Table 4.1.

The excess demand functions for all assets are now found by adding the demand functions of all agents for that asset, and subtracting the supply which comes from the agents' initial endowments. We obtain

$$\begin{aligned} z_j(\pi_1, \ldots, \pi_n) &= \sum_{i=1}^k x_i^j(\pi_1, \ldots, \pi_n) - \sum_{i=1}^k \omega_i^j \\ &= \sum_{i=1}^k \frac{p_j^i}{\pi_j} \sum_{\ell=1}^n \pi_\ell \omega_i^\ell - \sum_{i=1}^k \omega_i^j. \end{aligned}$$

Competitive equilibrium for log utility agents under objective probabilities

Inputs:

$k =$ number of agents

$n =$ number of states

$p_j \quad (j = 1, \dots, n)$ objective probabilities

$\omega_i^j \quad (i = 1, \dots, k;\, j = 1, \dots, n)$ initial endowments

Define:

$\omega_{\mathrm{a}}^j = \sum_{i=1}^{k} \omega_i^j$ aggregate payoff

Outputs:

risk-adjusted probabilities

$$q_j \propto \frac{1}{\omega_{\mathrm{a}}^j}\, p_j$$

equilibrium allocation

$$x_i^j = \frac{E^Q \omega_i}{E^Q \omega_{\mathrm{a}}}\, \omega_{\mathrm{a}}^j$$

Table 4.1: Competitive equilibrium allocation for log utility agents. The proportionality constant for the risk-adjusted probabilities is determined by the requirement that the probabilities must add up to 1.

The equilibrium prices are obtained by setting all of these functions equal to zero. The resulting equations can be written as

$$\sum_{\ell=1}^{n}\Big(\sum_{i=1}^{k} p_j^i \omega_i^\ell\Big)\pi_\ell = \Big(\sum_{i=1}^{k}\omega_i^j\Big)\pi_j \qquad (j=1,\dots,n). \tag{4.5}$$

This is a set of n homogeneous *linear* equations in the n unknowns $\pi_1,\dots,\pi_n$. It is a consequence of Walras' law that the equations are dependent, so that there exists a nonzero solution.

Let us consider the particular case in which there are only two assets and only two agents. For clarity of notation, the two agents will be referred to as A and B rather than as agent 1 and agent 2. The equations (4.5) become

$$\begin{aligned}(p_1^A\omega_A^1 + p_1^B\omega_B^1)\pi_1 + (p_1^A\omega_A^2 + p_1^B\omega_B^2)\pi_2 &= (\omega_A^1+\omega_B^1)\pi_1\\ (p_2^A\omega_A^1 + p_2^B\omega_B^1)\pi_1 + (p_2^A\omega_A^2 + p_2^B\omega_B^2)\pi_2 &= (\omega_A^2+\omega_B^2)\pi_2.\end{aligned}$$

The two equations are actually the same, as is seen from the fact that both equations can be rewritten as

$$(p_2^A\omega_A^1 + p_2^B\omega_B^1)\pi_1 - (p_1^A\omega_A^2 + p_1^B\omega_B^2)\pi_2 = 0.$$

To derive this in the case of each equation, set the difference of the left hand side and the right hand side equal to zero and recall that p_i^j denotes a probability, so that $1-p_1^A = p_2^A$ and $1-p_1^B = p_2^B$. This is just a manifestation of the general fact, discussed above, that there is a dependency in the equations of general equilibrium. We arrive at the following conclusion for a single-period market with two assets and two agents who have logarithmic utilities: prices π_1 and π_2 constitute an equilibrium price system if and only if

$$\frac{\pi_1}{\pi_2} = \frac{p_1^A\omega_A^2 + p_1^B\omega_B^2}{p_2^A\omega_A^1 + p_2^B\omega_B^1}\,. \tag{4.6}$$

The prices are determined up to a proportionality factor. The corresponding risk-adjusted probabilities are

$$q_1 = \frac{p_1^A\omega_A^2 + p_1^B\omega_B^2}{p_1^A\omega_A^2 + p_2^A\omega_A^1 + p_1^B\omega_B^2 + p_2^B\omega_B^1} \tag{4.7a}$$

$$q_2 = \frac{p_2^A\omega_A^1 + p_2^B\omega_B^1}{p_1^A\omega_A^2 + p_2^A\omega_A^1 + p_1^B\omega_B^2 + p_2^B\omega_B^1}\,. \tag{4.7b}$$

Once these probabilities have been computed, the equilibrium allocation can be obtained from (4.4). The results of the computations are illustrated for a number of specific cases in the examples below.

Example 4.1.1 We return to Example 3.1.1 again. Suppose that ice cream seller Alice and cinema owner Bob both have logarithmic utilities and that they both believe that the

	ω^{S}	ω^{R}	p_{S}	p_{R}	x^{S}	x^{R}	q_{S}	q_{R}
A	20	5	0.7	0.3	16.25	10.83	0.6087	0.3913
B	10	15			13.75	9.17		

Table 4.2: Data and equilibrium outcomes for Example 4.1.1.

probability of a sunny day is 0.7. The preference level associated to a portfolio consisting of x^{S} units of state contract S and x^{R} units of asset R is then given by

$$U_i(x_i^{\mathrm{S}}, x_i^{\mathrm{R}}) = 0.7 \ln x_i^{\mathrm{S}} + 0.3 \ln x_i^{\mathrm{R}}$$

where i is A for Alice or B for Bob.[1] Table 4.1.1 shows the equilibrium results. The equilibrium contract states that Alice will transfer 3.75 units to Bob on a sunny day, whereas Bob will transfer 5.83 units to Alice when the day is rainy. As predicted by the theory above, the allocation rule is proportional: whether the day is rainy or the sun shines, Alice gets 54% of aggregate income and Bob gets 46%. The agreement is beneficial for both of them, as can be verified by computing the preference levels with and without the contract:

$$\begin{aligned}
U_A(20, 5) &= 0.7 \ln 20 + 0.3 \ln 5 = 2.5798 \\
U_A(16.25, 10.83) &= 0.7 \ln 16.25 + 0.3 \ln 10.83 = 2.6665 \\
U_B(10, 15) &= 0.7 \ln 10 + 0.3 \ln 15 = 2.4242 \\
U_B(13.75, 9.17) &= 0.7 \ln 13.75 + 0.3 \ln 9.17 = 2.4994.
\end{aligned}$$

This shows that the equilibrium allocation is indeed better both for Alice and for Bob, but how big the improvement is cannot be read off from the figures above, since there is no calibration of the scale for preferences.

The size of the improvement brought about by the contract can be made tangible for instance introducing the notion of *acceptable cost*. For Alice, this is the number y that solves the equation

$$U_A(20, 5) = U_A(16.25 - y, 10.83 - y).$$

The acceptable cost can be interpreted as the maximum amount that Alice would be willing to pay to an intermediary who can establish the equilibrium contract. The acceptable cost for Alice is 1.17 units, and for Bob it is 0.86 units. Since an intermediary can charge both Alice and Bob for the services provided, one might say that the earning potential for

[1] This specification means in particular that the two agents only derive utility from their own income, not from each other's income. Apparently Alice and Bob do not have any special feelings for each other, either positive or negative.

financial intermediation in the situation of our example is 2.03 units, representing 7.8% of total market value (at the equilibrium prices).[2]

In the example we see that the adjusted probability of a sunny day is lower than the actual probability according to Alice and Bob, whereas the adjusted probability of a rainy day is higher than the corresponding actual probability. This is in line with the fact, as discussed above, that the income of Alice and Bob together is higher on sunny days than it is on rainy days. From the point of view of his initial endowment, Bob would hope for bad weather, but after the contract has been signed he earns more on sunny days, just like Alice. This reflects the comonotonicity property that was mentioned above.

One can verify from the data in Table 4.2 that all conditions for equilibrium are satisfied. Market clearing is evident from the fact that the entries in the column for x^{S} add up to the same number as the entries under ω^{S}, and the same relation holds for the entries in the columns for x^{R} and ω^{R}. Financial fairness means that, under the state prices as indicated by the adjusted probabilities, the 3.75 units of state contract S given by Alice to Bob are worth the same as the 5.83 units of state contract R that are given by Bob to Alice. Indeed we have $3.75 \cdot 0.6087 = 2.28 = 5.83 \cdot 0.3913$. Finally, to verify optimality, we can check that the proportionality property (3.13) is satisfied. The vectors of marginal utilities at the equilibrium allocations for Alice and Bob and the vector of ratios of the q-probabilities with respect to the p-probabilities are given by

$$\begin{bmatrix} 1/x_A^{\mathrm{S}} \\ 1/x_A^{\mathrm{R}} \end{bmatrix} = \begin{bmatrix} 0.0615 \\ 0.0923 \end{bmatrix}, \quad \begin{bmatrix} 1/x_B^{\mathrm{S}} \\ 1/x_B^{\mathrm{R}} \end{bmatrix} = \begin{bmatrix} 0.0727 \\ 0.1091 \end{bmatrix}, \quad \begin{bmatrix} q_{\mathrm{S}}/p_{\mathrm{S}} \\ q_{\mathrm{R}}/p_{\mathrm{R}} \end{bmatrix} = \begin{bmatrix} 0.8696 \\ 1.3043 \end{bmatrix}.$$

It can be verified by direct computation that these three vectors are proportional to each other.

Example 4.1.2 Consider the same situation as in the previous example, but now assume that the probabilities of sunshine and rain are 0.8 and 0.2 respectively. What effect does this have on the equilibrium outcomes? The results of computations are shown in Table 4.3. The equilibrium allocation is now closer to the initial endowments, which indicates that there is less scope for trading. This is shown also in the earning potential of financial intermediation, which is 1.64 units as opposed to 2.03 units in the previous example. These phenomena may be explained by noting that there is less uncertainty in the economy compared to the situation of Example 4.1.1.

Example 4.1.3 It is also of interest to consider the situation in which Alice and Bob do not agree on the probabilities of sunshine and rain. Suppose that Alice believes that the probability of a sunny day is 0.8, whereas Bob thinks it is only 0.6. The equilibrium outcomes are shown in Table 4.4. Only the subjective probabilities of Alice and Bob play a role

[2] Here it is assumed that the intermediary will establish the equilibrium contract. Some alternative assumptions are explored in Exercises 3.6.4 and 3.6.5.

	ω^{S}	ω^{R}	p_{S}	p_{R}	x^{S}	x^{R}	q_{S}	q_{R}
A	20	5	0.8	0.2	17.50	11.67	0.7273	0.2727
B	10	15			12.50	8.33		

Table 4.3: Data and equilibrium outcomes for Example 4.1.2.

	ω^{S}	ω^{R}	p_{S}	p_{R}	x^{S}	x^{R}	q_{S}	q_{R}
A	20	5	0.8	0.2	18.46	7.50	0.6190	0.3810
B	10	15	0.6	0.4	11.54	12.50		

Table 4.4: Data and equilibrium outcomes for Example 4.1.3.

in the equilibrium results; it does not matter what the true ("objective") probabilities are. The fact that the agents' probability assessment deviate from each other in the optimistic direction for both of them does not encourage trading, as is seen from the results. The limited scope for trading is also reflected in the earning potential for financial intermediation, which is only 0.39 units, or 1.5% of total market value. Another effect which may be noted is that the comonotonicity property does not apply. Recall that this property was derived above under the assumption that the agents agree on the probabilities of events. If there is disagreement, the property may not hold as is seen in the present example.

Example 4.1.4 Suppose now that rainy days are not so bad, and that Alice and Bob earn 10 and 20 units respectively when it rains, instead of 5 and 15 as assumed before. The total earnings of Alice and Bob are then the same on a sunny day and on a rainy day. Suppose also that Alice and Bob agree that the probability of a sunny day is 0.7. The equilibrium outcomes for this case are shown in Table 4.5. The table shows some remarkable features: the q-probabilities are the same as the p-probabilities, and in the equilibrium allocation both Alice and Bob become indifferent, at least from a financial point of view, as to whether it rains or the sun is shining. These facts are a consequence of the principle that, when there is no aggregate risk, the equilibrium allocation eliminates individual risk. This principle holds under the assumption that the agents agree on the probabilities of the states. The next example considers what happens if there is disagreement.

Example 4.1.5 Take the same situation as in the previous example, but assume now that the probability assessments of Alice and Bob are as in Example 4.1.3. The equilibrium outcomes are given in Table 4.6. It is seen that the equilibrium allocation is no longer riskfree.

	ω^{S}	ω^{R}	p_{S}	p_{R}	x^{S}	x^{R}	q_{S}	q_{R}
A	20	10	0.7	0.3	17	17	0.7	0.3
B	10	20			13	13		

Table 4.5: Data and equilibrium outcomes for Example 4.1.4.

	ω^{S}	ω^{R}	p_{S}	p_{R}	x^{S}	x^{R}	q_{S}	q_{R}
A	20	10	0.8	0.2	19.20	12	0.7143	0.2857
B	10	20	0.6	0.4	10.80	18		

Table 4.6: Data and equilibrium outcomes for Example 4.1.5.

4.2 Exponential utility

Again we consider a market for single-period financial assets, but now we assume that all agents have *exponential* utility functions. In this section we always assume that the agents agree on the probabilities of the states. On the other hand, while in the previous section all agents had the same risk aversion expressed by the logarithmic utility function, we now allow agents to have different levels of risk aversion. Specifically, the preference function of agent i is given by

$$U_i(x_i^1, \dots, x_i^n) = -\tau_i \sum_{j=1}^{n} p_j e^{-x_i^j/\tau_i}.$$

The quantity τ_i indicates the risk tolerance of agent i. One may perhaps think of τ_i as an amount so large that a loss of this size would make agent i have a bad night of sleep. Such amounts can indeed be different for different people.

Again we follow the procedure of Section 3.3. The optimality conditions in this case are

$$p_j e^{-x_i^j/\tau_i} = \lambda_i \pi_j.$$

From this we obtain

$$x_i^j = -\tau_i \ln \frac{\lambda_i \pi_j}{p_j}\,. \tag{4.8}$$

The market clearing equations are

$$-\sum_{i=1}^{k} \tau_i \ln \frac{\lambda_i \pi_j}{p_j} = \sum_{i=1}^{k} \omega_i^j \qquad (j = 1, \dots, n).$$

This can be rewritten as

$$-\sum_{i=1}^{k} \tau_i \ln \lambda_i - \ln \frac{\pi_j}{p_j} \sum_{i=1}^{k} \tau_i = \omega_{\mathrm{a}}^j \qquad (j = 1, \dots, n)$$

where $\omega_{\mathrm{a}}^j := \sum_{i=1}^k \omega_i^j$ denotes the aggregate payoff in state j. To simplify the notation further, also write $\tau_{\mathrm{a}} := \sum_{i=1}^k \tau_i$. We then find

$$\ln \frac{\pi_j}{p_j} = -\frac{\omega_{\mathrm{a}}^j}{\tau_{\mathrm{a}}} + \sum_{i=1}^k \frac{\tau_i}{\tau_{\mathrm{a}}} \ln \lambda_i \tag{4.9}$$

and hence

$$\pi_j = p_j e^{-\omega_{\mathrm{a}}^j/\tau_{\mathrm{a}}} \exp\Big(\sum_{i=1}^k \frac{\tau_i}{\tau_{\mathrm{a}}} \ln \lambda_i\Big).$$

Since the exponential factor on the right hand side does not depend on j, we can now already determine the adjusted probabilities:

$$q_j = \frac{p_j e^{-\omega_{\mathrm{a}}^j/\tau_{\mathrm{a}}}}{\sum_{\ell=1}^n p_\ell e^{-\omega_{\mathrm{a}}^\ell/\tau_{\mathrm{a}}}}\,. \tag{4.10}$$

The expression (4.8) for the equilibrium allocation can be rewritten as

$$x_i^j = -\tau_i \ln \lambda_i - \tau_i \ln \frac{\pi_j}{p_j}\,.$$

Using (4.9), we see that the equilibrium payoffs are of the form

$$x_i^j = \frac{\tau_i}{\tau_{\mathrm{a}}} \omega_{\mathrm{a}}^j + c_i \tag{4.11}$$

where the c_i's are numbers that depend on the index i but not on j. It is convenient to replace the Lagrange multipliers λ_i as unknowns by the constants c_i. The c_i's are determined from the budget equations. Noting that these equations may be formulated in terms of the normalized state prices q_j, we can write

$$\sum_{j=1}^n q_j \Big(\frac{\tau_i}{\tau_{\mathrm{a}}} \omega_{\mathrm{a}}^j + c_i\Big) = \sum_{j=1}^n q_j \omega_i^j$$

or in other words

$$\frac{\tau_i}{\tau_{\mathrm{a}}} \sum_{j=1}^n q_j \omega_{\mathrm{a}}^j \; + c_i = \sum_{j=1}^n q_j \omega_i^j$$

and hence

$$c_i = \sum_{j=1}^n q_j \omega_i^j - \frac{\tau_i}{\tau_{\mathrm{a}}} \sum_{j=1}^n q_j \omega_{\mathrm{a}}^j.$$

Using the interpretation of the q_j's as probabilities, this formula can also be written in a more compact form as

$$c_i = E^Q \omega_i - \frac{\tau_i}{\tau_{\mathrm{a}}} E^Q \omega_{\mathrm{a}}.$$

Taking this together with (4.11), we obtain

$$x_i^j = E^Q \omega_i + \frac{\tau_i}{\tau_{\mathrm{a}}} \big(\omega_{\mathrm{a}}^j - E^Q \omega_{\mathrm{a}}\big). \tag{4.12}$$

It is readily verified from this expression that both the budget constraints (which may be written as $E^Q x_i = E^Q \omega_i$) and the market clearing conditions $\sum_{i=1}^k x_i^j = \omega_{\mathrm{a}}^j$ are satisfied.

In words, the result may be formulated as follows: in equilibrium, the payoff for agent i in state j consists of the average payoff of the agent across all states, plus a fixed percentage of the difference between the aggregate payoff in state j and the average aggregate payoff. The averages mentioned here are weighted averages, where the weights are given by the adjusted probabilities. It can be concluded that the equilibrium allocation in the case of exponential utility functions and agreement on probabilities is such that each agent takes a constant share of aggregate risk. Which percentage of aggregate risk is borne by agent i is determined by the ratio of τ_i, the agent's risk tolerance, to aggregate risk tolerance τ_{a}. It may also be noted that under the exponential utilities the payoff for agent i in state j is equal to the average payoff *plus* a state-dependent correction, whereas in the case of logarithmic utility the payoff is equal to the average payoff *times* a state-dependent factor (see (4.4)). The allocation rule for exponential utility agents is summarized in Table 4.7.

Example 4.2.1 Tables 4.8–4.11 show equilibria for four different market situations under exponential utility. In the first table, both Alice and Bob have a risk tolerance of 10 units of currency. They gain substantial benefits from trading; the equivalent wealth for Alice is 8% of her expected income, and for Bob it is even almost 10%. When the risk tolerance of both Alice and Bob is reduced to 5 units, these numbers increase to 13% and 20% respectively. The strong risk aversion of Alice and Bob is also reflected in the fact that the q-probabilities deviate quite a bit from the p-probabilities. In the situation of Table 4.8, the earning potential of financial intermediation is already substantial at 2.34 units, which represents 9% of total market value; under the stronger risk aversion in Table 4.9, the earning potential is even 4.32 units, representing 17.6% of total market value.

We can also consider situations in which Alice and Bob have different risk tolerances. Table 4.10 shows a situation in which Alice has a risk tolerance of 10 units, while Bob has a risk tolerance of 5 units so that he is more risk averse than Alice. In this situation, the utility of trading for Alice is decreased with respect to the situation in which her risk tolerance is also 5 units. The earning potential of financial intermediation is in this case 2.35 units, representing 9.2% of total market value. Finally, we consider a situation in which Alice and Bob still have different risk tolerances, but the earnings on rainy days are improved so that the aggregate income is now the same when it rains and when the sun shines, just as in Example 4.1.4. The equilibrium is now exactly the same as in that example, even though the utility functions are different. In fact, when there is no aggregate risk and there is agreement on probabilities, the equilibrium does not depend on the utility functions as long as agents are risk averse.

Competitive equilibrium for exponential utility agents under objective probabilities

Inputs:

$k =$ number of agents

$n =$ number of states

$p_j \quad (j = 1, \ldots, n)$ objective probabilities

$\tau_i \quad (i = 1, \ldots, k)$ risk tolerances

$\omega_i^j \quad (i = 1, \ldots, k; j = 1, \ldots, n)$ initial endowments

Define:

$\tau_{\text{a}} = \sum_{i=1}^{k} \tau_i$ aggregate risk tolerance

$\omega_{\text{a}}^j = \sum_{i=1}^{k} \omega_i^j$ aggregate payoff

Outputs:

risk-adjusted probabilities

$$q_j \propto e^{-\omega_{\text{a}}^j/\tau_{\text{a}}}\, p_j$$

equilibrium allocation

$$x_i^j = E^Q \omega_i + \frac{\tau_i}{\tau_{\text{a}}}\big(\omega_{\text{a}}^j - E^Q \omega_{\text{a}}\big)$$

Table 4.7: Competitive equilibrium allocation for exponential utility agents. The proportionality constant for the risk-adjusted probabilities is determined by the requirement that the probabilities must add up to 1.

	ω^{S}	ω^{R}	τ	p_{S}	p_{R}	x^{S}	x^{R}	q_{S}	q_{R}
A	20	5	10	0.7	0.3	15.86	10.86	0.5860	0.4140
B	10	15	10			14.14	9.14		

Table 4.8: Equilibrium under exponential utility. Inputs on the left, outputs on the right.

	ω^{S}	ω^{R}	τ	p_{S}	p_{R}	x^{S}	x^{R}	q_{S}	q_{R}
A	20	5	5	0.7	0.3	14.62	9.62	0.4619	0.5381
B	10	15	5			15.38	10.38		

Table 4.9: Equilibrium under exponential utility. Inputs on the left, outputs on the right.

	ω^{S}	ω^{R}	τ	p_{S}	p_{R}	x^{S}	x^{R}	q_{S}	q_{R}
A	20	5	10	0.7	0.3	16.21	9.54	0.5450	0.4550
B	10	15	5			13.79	10.46		

Table 4.10: Equilibrium under exponential utility. Inputs on the left, outputs on the right.

	ω^{S}	ω^{R}	τ	p_{S}	p_{R}	x^{S}	x^{R}	q_{S}	q_{R}
A	20	10	10	0.7	0.3	17	17	0.7	0.3
B	10	20	5			13	13		

Table 4.11: Equilibrium under exponential utility; situation of no aggregate risk.

4.3 Exercises

1. In a situation in which there are two log utility agents with subjective probabilities, and two events that may take place, analyze the equilibrium when it is assumed that the initial endowment of one of the agents is in both events much larger than the endowment of the other agent. To do this, let ω_B^1 and ω_B^2 be fixed, and determine the limits of q_1 and q_2 as given by (4.7) when $\omega_A^1 \to \infty$, $\omega_A^2 \to \infty$, and

$$\lim \frac{\omega_A^1}{\omega_A^2} = \alpha$$

for some $\alpha > 0$. Do the subjective probabilities of agent B matter for the result? Also consider the special case in which the rich agent A has the same income in both events, or in other words, $\alpha = 1$. Find the equilibrium portfolio of agent B in the limit. Do the subjective probabilities of agent B play a role in the composition of this portfolio? What does the portfolio of agent B look like if both agents agree on the probabilities of the events?

2. Consider a situation as in Section 4.2, with k agents whose preference functions are given by exponential utility. Analyze the equilibrium when it is assumed that one of the agents is very risk tolerant. To do this, let $\tau_2, \ldots, \tau_k$ be fixed, and determine the limits of the risk-adjusted probabilities q_j as τ_1 tends to infinity. Also find the limits of the equilibrium portfolios of all agents.

3. Consider a single-period economy in which the agents all use power utility (see Table 2.1) with the same risk aversion coefficient γ. The agents also use the same objective probabilities $p_1, \ldots, p_n$. The initial endowment of agent i is given by $(\omega_i^1, \ldots, \omega_i^n)$. The total payoff when outcome j occurs ("aggregate income in state j") is denoted by ω_{a}^j; in other words, we have

$$\omega_{\mathrm{a}}^j = \sum_{i=1}^{k} \omega_i^j.$$

Define numbers q_j (which will act as adjusted probabilities) by

$$q_j \propto (\omega_{\mathrm{a}}^j)^{-\gamma} p_j \tag{4.13}$$

where the constant of proportionality is determined by the requirement that the sum of the q_j's should be equal to 1.[3] Define an allocation as follows:

$$x_i^j = \frac{E^Q \omega_i}{E^Q \omega_{\mathrm{a}}} \, \omega_{\mathrm{a}}^j \tag{4.14}$$

where by definition

$$E^Q \omega_i = \sum_{j=1}^{n} q_j \omega_i^j \quad (i = 1, \ldots, k), \qquad E^Q \omega_{\mathrm{a}} = \sum_{j=1}^{n} q_j \omega_{\mathrm{a}}^j.$$

[3] In other words, the definition above means that $q_j = \frac{1}{\alpha} (\omega_{\mathrm{a}}^j)^{-\gamma} p_j$, where $\alpha = \sum_{j=1}^{n} (\omega_{\mathrm{a}}^j)^{-\gamma} p_j$.

Show directly from the equilibrium equations that the proposed allocation constitutes a competitive equilibrium. (Note that the allocation rule is proportional, just as in the case of log utility. In fact, the case of log utility can be viewed as a limit case of the above as the risk aversion coefficient γ tends to 1.)

4. Consider an economy in which the normalized pricing kernel and the return on the market portfolio are related by

$$\theta \propto (R^m)^{-\gamma}$$

where γ is a positive constant. (The previous exercise provides an example; note that the proportionality constant is determined by the requirement $E[\theta] = 1$ once the distribution of R^m is given.) Assume that R^m follows a lognormal distribution:

$$R^m = \exp(\mu_m + \sigma_m Z_m), \qquad Z_m \sim N(0,1).$$

Now consider an asset whose return R is lognormally distributed as well:[4]

$$R = \exp(\mu + \sigma Z), \qquad Z \sim N(0,1).$$

Also assume that Z and Z_m are jointly normally distributed with correlation coefficient ρ. In this setting, it is convenient to represent the riskfree return as $R^f = e^r$. For the computations below, recall the standard formula for the expectation of a lognormal variable:

$$E[\exp(\mu + \sigma Z)] = \exp(\mu + \tfrac{1}{2}\sigma^2) \qquad (Z \sim N(0,1)).$$

a. Find the relation between the parameters that follows from the equality $E[\theta R] = R^f$.

b. Find the relation between the parameters that follows from the equality $E[\theta R^m] = R^f$.

c. Determine $E[R^e]/E[R^{me}]$, and compare the result to the CAPM beta.

[4]Note that, in these parametrizations, the parameters μ_m and μ should be looked at as *net* return parameters (typically small positive numbers) rather than as *gross* return parameters (typically numbers a bit larger than 1).

Chapter 5

Consistent pricing

In the previous chapter, prices of assets were derived from assumptions on the preference functions of agents and from a particular notion of equilibrium. This type of analysis leads to insights concerning the way in which asset prices depend on agents' exposure to risk and on their risk preferences. These insights are useful to provide a broad explanation of certain phenomena observed in financial markets. To give practical support for pricing in daily market practice, however, the model of the previous chapter is idealized to a too great extent. The present chapter deals with another approach which is at the same time less ambitious and more ambitious than the approach of the previous chapter. It is less ambitious, in that it does not attempt to explain prices of all assets, but only of what might be called derivative assets; prices of "basic assets" are taken as *inputs* of the model, rather than as outputs. It is more ambitious because it aims for much greater accuracy of pricing than one can ever hope to achieve on the basis of the equilibrium approach. The equilibrium theory of the previous chapter and the arbitrage theory of this chapter lead to pricing formulas of a similar form, but they have different ways of motivating the parameters that appear in this form. The two theories should be thought of as complementary, rather than as antagonistic.

5.1 The notion of arbitrage

English-Dutch vocabulary for Section 5.1	
arbitrage arbitrage, onmiddellijke winst	**arbitrage opportunity** arbitragemogelijkheid
arbitrage-free arbitragevrij	

The notion of *arbitrage* may be defined as follows. The definition is stated here for the case of financial assets so that we can speak of cashflows, but a similar definition might also be formulated for physical assets.

Definition 5.1.1 An *arbitrage* is a trading strategy that, with positive probability, generates a positive cashflow either immediately or at some point in the future, and that does not generate any negative cashflows.

The sign convention that is used here is that a positive cashflow refers to an amount of money that is *received*, while a negative cashflow indicates an amount that is *paid*. The statement that the strategy does not generate any negative cashflows means in particular that no initial investment is needed. The definition may therefore be rephrased briefly by saying that "an arbitrage creates something out of nothing." A related but somewhat stronger notion is the following.

Definition 5.1.2 A *strict arbitrage* is a trading strategy that generates a positive cashflow immediately, and that does not generate any negative cashflows.

Just like an arbitrage, a *strict* arbitrage creates money out of nothing, because a positive cashflow is received which is not counterbalanced by any liability in the future. The notion of strict arbitrage is stronger however because it requires that the positive cashflow is available *immediately*, rather than possibly later, and with *certainty*, rather than only with positive probability.

The notion of "probability" has to interpreted with some care here, because agents may not agree about the probability of events. In the definition above however, only the notion of "positive probability" is used. Therefore the underlying assumption is not that agents agree on the probabilities of all events, but that there is agreement on which events have positive probability, that is, which events are possible and which are not. Also the notion of "trading strategy" has to be interpreted with care, as will be demonstrated in the examples below.

In practice, there is not much difference between arbitrage and strict arbitrage. Even if a positive cashflow takes place in the future and depends on an event that is not sure to happen, as long as agents agree that there is a possibility that the event will happen, the rights to the cashflow do have a positive market price. Therefore an uncertain profit in the future can be turned into a certain profit now. Of course, the market price will include a discount which will be stronger when the cashflow takes place further into the future and the event on which it is based has a lower (perceived) probability.

The *principle of absence of arbitrage* states that market prices must be such that there are no arbitrage opportunities. Analogously one can formulate a principle of absence of strict arbitrage. Whether there exist arbitrage opportunities in actual financial markets is a matter of perspective. In fact, some financial institutions employ so-called *arbitrageurs* whose job it is to detect and utilize temporary small imbalances for instance in exchange rates between currencies. The fact that there are people who make a living in this way proves that arbitrage opportunities must exist. At the same time, for anyone who is not in the arbitrage business, the world appears as if there are no arbitrage opportunities, because the effect of the actions of the arbitrageurs is to eliminate quickly any such opportunities whenever they arise. Therefore, mathematical models of financial markets, unless they are specifically intended for the exploitation of arbitrage opportunities, are constructed in such a way that they are free of arbitrage.

A few examples relating to the notion of arbitrage, both in deterministic and stochastic cases, are given below.

Example 5.1.3 Alice can get a one-year loan from Easyloan Bank at a rate of 5%. She can make a one-year deposit at Supersave Bank at a rate of 5.2%. Under the assumption that Supersave will certainly not go into default during the next year, Alice has an arbitrage opportunity. She even has a strict arbitrage opportunity.

Example 5.1.4 The Wimbledon final is played between Venus and Serena. Bob is willing to bet 3:1 that Venus will win; that is, he is willing to sign a contract with any counterparty under which, immediately after the final has been played, Bob will pay to the counterparty three euros if Serena has won, and the counterparty will pay to Bob one euro if Venus is the winner. On the other hand, Carol is willing to bet 3:1 that Serena will win. Now Daniel has an arbitrage opportunity and he can make money from the tennis tournament, even though all that he knows about tennis is that at the end one of the players has won and the other has lost.

Example 5.1.5 Suppose you can change US dollars to British pounds at a rate of 2.42 dollars to the pound, pounds to euros at a rate of 0.59 pounds to the euro, and euros to dollars at a rate of 0.70 euros to the dollar. You have an arbitrage opportunity.

Example 5.1.6 When you put 100 euro in a savings account that carries a 4% interest rate (discretely compounded), then after one year you have 104 euro. In a sense you have made a riskless profit of 4 euro. However this is *not* an arbitrage, because the initial investment is nonzero. To get the 4 euro, you needed to keep the 100 euro in the savings account all the time, so that you could not use it for consumption. Moreover you are subject to inflation risk; if inflation is high, it might even happen that 104 euro will buy you less next year than you can get for 100 euro now.

Example 5.1.7 Suppose that by paying 1 euro you can obtain 99% probability of gaining 10 000 euro immediately, and 1% probability of gaining 95 cents. This is an excellent investment opportunity but *not* an arbitrage, since there is a positive probability of a loss of five cents.

Example 5.1.8 Consider a simple coin tossing game in which you have 50% probability of losing your stake and 50% probability of doubling it. You can choose how much you want to bet, and there is an unlimited number of rounds. Payment is only settled when you decide to stop playing the game. You may apply a strategy as follows. Suppose that you start with nothing. In the first round, bet one euro. If you win, you get two euros, and after settling the account there is one euro left; in this case, stop the game. If you lose, you are one euro in debt. In that case, continue the game and bet two euros in the next round. If you win the toss this time, your cumulative profit is one euro. If you lose, your debt is three euros. In the latter case, continue the game and bet four euros in the next round. Keep doubling the stakes every time you lose the toss, and quit the game as soon as you win a toss. The probability of a negative outcome is less than the probability of losing n successive tosses, which is 2^{-n}, for any finite number n; in other words, the probability of a negative outcome is zero. So, by playing this strategy, it seems you can start with nothing and end up with one euro for sure.[1] Note that, theoretically speaking, the game does not need to take an unlimited amount of time; one can assume that the second round is played one minute after the first round, the third round is played thirty seconds later, the fourth round follows after fifteen seconds, and so on; the game will then last no more than two minutes.

In the setting in which the number of rounds you can play is unlimited and the amounts that you can bet are also unlimited, the coin tossing game does indeed allow an arbitrage opportunity. As soon as an upper bound is imposed on either the number of rounds that can be played or on the amounts that may be bet, the arbitrage opportunity disappears. This example shows that whether or not arbitrage exists in a given market also depends on the specification of trading strategies that may be applied. Therefore, when we proceed below to a mathematical definition of a market, the admissible strategies will be part of the definition.

5.2 Example: defaultable bonds

Governments as well as corporations and individuals often lend money to finance their operations, and frequently loans need to be renewed. As a consequence there is a large and lively market for “sovereigns” (government bonds), “commercial paper” (short-term corporate bonds), “junk bonds” (bonds from issuers with weak credit status), and many other forms of debt. A characteristic of loans, known since ancient times, is that there is a risk that the borrower will not pay back the money lent at the agreed time, or may pay back only part of it. This is called the *default risk*. Naturally, this risk has an impact on prices. Indeed, if bonds issued by borrowers that enjoy a solid reputation would be available at the same price as bonds of dubious debtors, then everybody would buy the first and not the second. The bonds with lower credit status must have a lower price. Equivalently, weak debtors must promise to pay a higher interest rate on their loans than strong debtors do; the higher rate compensates for the default risk. This is why junk bonds are also known, in particular in advertisements aimed towards potential buyers, as “high-yield bonds”. The

[1] To make it more interesting, you might also start by betting one million euro in the first round, and then follow the same strategy of doubling the bet after every loss. In this way you end up with one million euro for sure.

difference between the actual rate paid by an issuer and the “riskless” rate (the one paid by the most solid issuers) is known as the *credit spread*.

For investors who want to invest part of their money in bonds, it is important to have information about the credit status of issuers. Such information is provided on a systematic basis by *rating agencies*. These agencies categorize national governments, corporations, and other major issuers according to their probability of default as estimated by the analysts who work for the agencies. Several systems are in use, but most common is the classification in which the most reliable issuers are in category AAA (“triple-A”), and the next categories in order of increasing probability of default are AA, A, BBB, BB, and so on until C, with an additional category D for issuers who already behind on their debt service.[2] The credit rating of a particular issuer can be used by investors to judge whether the yield that is offered is sufficiently attractive.

While there is a liquid secondary market for government bonds and bonds issued by large corporations, it would be difficult to set up a similar market for, say, individual mortgages or loans to small companies. For this reason, such loans are often packages into large bundles which may contain, for instance, tens of thousands of mortgages. After a bundle has been formed, it may be split into so-called *tranches* which each have their own credit status. Each tranche corresponds to what might be called a layer of defense against payment defaults. The lowest-rated (or even unrated) tranche is the one that picks up the first losses. If losses exceed the absorbing capacity of the first tranche, then the second tranche comes into play, and so on. Since there are many individual debtors in the bundle, it may be expected on the basis of the law of large numbers that the upper tranches are quite safe and can have a higher credit status than any of the individual debtors would have. In this way, money from institutional investors with a policy of investing only in safe assets can be used to support a large part of funding that goes to end users with a low credit status. The debt instruments that are created in this way are called *collateralized debt obligations* (CDOs).

In the early years of the 21st century, CDOs became very popular in the US, in particular in the market for *mortgage-backed securities*, and many believe that the possibilities unleashed by CDOs were a main contributing factor to the subprime mortgage crisis of 2007. To see the way in which CDOs can be used for what is sometimes called “rating arbitrage”, let us consider a highly simplified example which nevertheless does represent the main characteristics of the situation.

Suppose we have two bonds that have a maturity of one year, in other words, loans that should be paid back one year from now. The bonds are from different issuers, which are both not in terribly good shape; let us say that for each of the two issuers there is a 10% probability of default. Usually, when a default occurs it doesn’t mean that none of the money is paid back, but for instance that only half of the money is paid back. In

[2] Additional refinements are made by adding $+$ or $-$; for instance AA+ is used to indicate a category between AA and AAA.

this case we say that there is a 50% recovery rate. Suppose that the expected recovery for both of the issuers is 50%. Suppose the face value of the bond is 100; the expected amount that will be paid after one year is then equal to 95 for both issuers. Suppose that the discretely compounded riskfree rate for the maturity of one year is 3%. If we take the expected amount that will be paid back as a basis for pricing and we suppose that there will be no intermediate interest payments, then the price of the bond should be equal to $95/1.03 = 92.23$. Comparing this to the bond's face value 100, the yield (in case of no default) of the bond can be computed as $(100/92.23) - 1 = 0.0842$. In other words, the credit spread on this basis is 5.42 percentage points, that is, 542 basis points.[3] As argued in previous chapters, risk aversion that is present in the market may cause actual prices to deviate from the value that would be computed on the basis of statistical expectations. Since defaults are more likely to happen in bad states of the economy, the effect of risk aversion should be to increase the spread to somewhat more than 542 bps.

There are four situations that may arise at the end of the year: neither of the two issuers defaults; the first issuers defaults but the second does not; the second issuer defaults but the first does not; and finally, both issuers default. If we assume that the default events for the two issuers are independent, then the probabilities of these four states are, respectively, 81%, 9%, 9%, and 1%. The four possible outcomes for the bundle consisting of the two portfolios can be represented as follows:

$$\begin{array}{|c|c|}\hline 100 & 100 \\ \hline 50 & 50 \\ \hline \end{array} + \begin{array}{|c|c|}\hline 100 & 50 \\ \hline 100 & 50 \\ \hline \end{array} = \begin{array}{|c|c|}\hline 200 & 150 \\ \hline 150 & 100 \\ \hline \end{array} \tag{5.1}$$

Tranching means that the bundle is split up in a different way, without changing the total. For instance, an alternative division could be as follows:

$$\begin{array}{|c|c|}\hline 100 & 100 \\ \hline 100 & 100 \\ \hline \end{array} + \begin{array}{|c|c|}\hline 50 & 50 \\ \hline 50 & 0 \\ \hline \end{array} + \begin{array}{|c|c|}\hline 50 & 0 \\ \hline 0 & 0 \\ \hline \end{array} = \begin{array}{|c|c|}\hline 200 & 150 \\ \hline 150 & 100 \\ \hline \end{array} \tag{5.2}$$

Another alternative is given by

$$\begin{array}{|c|c|}\hline 150 & 150 \\ \hline 150 & 100 \\ \hline \end{array} + \begin{array}{|c|c|}\hline 50 & 0 \\ \hline 0 & 0 \\ \hline \end{array} = \begin{array}{|c|c|}\hline 200 & 150 \\ \hline 150 & 100 \\ \hline \end{array} \tag{5.3}$$

The leftmost tranche in (5.2) represents a riskless bond with face value 100. The tranche in the middle has the characteristics of a bond with face value 50 that has a default probability of 1% with 0% recovery rate. Finally, the rightmost tranche is a low-grade bond with face value 50, default probability 19%, and zero recovery. In (5.3), there are two tranches; one

[3] A *basis point* (bp) is 1/100th of a percentage point.

is the same low-grade bond as before, and the other represents a bond with face value 150, 1% default probability, and a 67% recovery rate. Clearly the tranches would get ratings that are different from the ratings of the original two bonds. Within the context of this simple example the usual ratings do not apply, but for the purpose of the discussion let us assume that there is a ratings table as shown in Table 5.1. The newly invented rating categories that appear here are RF (Risk Free), QR (Quite Reliable), RD (Rather Dubious), and JB (Junk Bond). Risk premia in this table are obtained by multiplying the default probabilities in each category by a number larger than 1. Such "multipliers" can also be obtained from real-world bond prices, by computing an implied probability of default as calculated from market prices and comparing this to the empirical probability of default for each rating category as listed by the rating agencies. In the real world, it is found that such multipliers are different for different rating categories, with multipliers for debtors with high credit status being somewhat larger than for debtors with low credit status. In line with this observation, multiplier 1.6 has been used in the table for category QR, 1.4 for category RD, and 1.2 for JB.

As an example of the calculations that lead to the numbers seen in Table 5.1, consider for instance the category QR. Given that the default probability is 1% and the recovery is zero, the expected repayment of principal is 99 if the bond's face value is 100. Discounting by the 3% fixed interest rate leads to the value $99/1.03 = 96.12$.[4] Therefore the spread without risk premium is found by solving for x such that $100/(1.03 + x) = 96.12$. One finds $x = 0.0104$; in other words the spread without risk premium is 104 basis points, as shown in the table. To calculate the spread with risk premium we multiply the default probability by the factor 1.6, so that the probability of default is now 1.6% rather than 1%. The corresponding expected repayment of principal is 98.4; from discounting we get $98.4/1.03 = 95.53$, which then is the market price of the bond with face value 100 in category QR. In order to find the corresponding credit spread, solve for x in $100/(1.03 + x) = 95.53$; we obtain $x = 0.0167$, or 167 bps.

At this point we can calculate the value of the portfolio in three different ways, based on the three subdivisions that appear in (5.1), (5.2), and (5.3). We find the following results:

$$2 * 90.29 = 180.58 \qquad (5.1)$$

$$97.09 \ + \ 0.5 * 95.53 \ + \ 0.5 * 74.95 = 182.33 \qquad (5.2)$$

$$1.5 * 96.57 \ + \ 0.5 * 74.95 = 182.34 \qquad (5.3)$$

It seems that tranching has increased the value of the portfolio. Since the total cashflows of the portfolio have not changed, the conclusion has to be that determination of prices in the way described above in fact leads to arbitrage opportunities. By buying separate risky debts and re-packaging them through a procedure of bundling and tranching, it is possible

[4]Numbers as shown in the text are rounded, but the actual successive calculations are carried out to machine accuracy.

default probability	recovery rate	rating category	spread without risk premium	spread with risk premium	price (face value 100)
0%	n.a.	RF	0	0	97.09
1%	67%	QR+	34	54	96.57
1%	0%	QR	104	167	95.53
10%	50%	RD	542	775	90.29
19%	0%	JB	2416	3042	74.95

Table 5.1: Hypothetical bond characteristics and bond values

to obtain profits by what is called *rating arbitrage*.

The size of the arbitrage opportunity may seem small in the example as worked out above, but the example included only two debtors; in bundles of thousands of debtors, the effect can be much stronger. On the other hand, it was assumed in the example that the defaults of the two debtors are independent events, whereas in most bundles it would be reasonable to assume that default events of different debtors are positively correlated. Positive correlation has a damping effect on rating arbitrage. Nevertheless, substantial effects can still be reached by including many debtors. Rating arbitrage presumably has played a substantial role in the spectacular growth and subsequent crash of the market for CDOs in the first decade of the 21st century.[5]

The pricing framework in the example is based on placing credit-sensitive assets into categories based on their associated default probabilities, and using a fixed spread for each category. We calculated the spreads for different rating categories on the basis of multiplication factors applied to the default probabilities, but since for any given rating category there is a one-to-one connection between multiplication factor and credit spread, we could also have prescribed the spreads directly. Many investment advisors have for many years based their assessments of bonds on the basis of the yields offered by these bonds in relation to their risk of default. Nevertheless, apparently something is wrong with this framework.

It can be verified that no arbitrage opportunities arise if default probabilities are not adjusted for credit risk, i.e., if prices are computed on the basis of expected payoffs, discounted at the riskfree rate. However, empirical research indicates that actual market prices are not formed in this way. The theory that has been discussed in the previous chapters also suggests that it is quite possible for equilibrium prices to differ from expected payoffs, and that

[5]The size of investments in CDOs in 2006 has been estimated at approximately $1.5 \cdot 10^{12}$ USD. In particular, about 700 billion USD worth of CDOs had been constructed on the basis of US housing mortgages, many of them in the "subprime" category (borrowers of dubious credit status). When US house prices declined after a peak in 2006, default probabilities needed to be revised. Massive downgrading of CDOs followed.

in fact such deviations should be expected in an economy populated by risk-averse agents. The challenge therefore is to construct pricing frameworks that are *internally consistent*, in the sense that they do not allow arbitrage opportunities, as well as *market consistent*, in the sense that the prices computed by the framework match prices that are actually observed in the market.[6] The construction of such consistent pricing frameworks will be the subject of the following chapters.

5.3 Exercises

1. Suppose that you can bet on the outcome of an event at a bookmaker. Assume that the event has two possible outcomes, which we call *up* and *down*. The odds of these two outcomes at the bookmaker are u and d respectively, which means that if you bet 1 euro on *up* and you have guessed right, you get u euros, and when you bet 1 euro on *down* and this outcome does appear, you get d euros. Of course, both u and d are larger than 1. A bookmaker is said to be "ideal" if $\frac{1}{u} + \frac{1}{d} = 1$. If this holds, the numbers $p = 1/u$ and $q = 1/d$ can be looked at as implied probabilities of the *up* and *down* events respectively. In this exercise, assume that discounting can be ignored (in other words, assume that you can borrow at zero interest rate).

a. Suppose you can in fact bet at two bookmakers, and that both bookmakers are ideal. Show that, if the odds offered by the two bookmakers are not exactly equal, you have an arbitrage opportunity, even under the restriction that you cannot bet a negative amount. (Hint: look for a combination of bets at the two bookmakers which is such that the total amount won is the same no matter whether *up* or *down* comes out.)

b. In a hypothetical future, the Champions League final is played between PSV and Willem II. There is a bookmaker in Eindhoven who will pay 3 euro for each euro that is placed on a victory of Willem II, and 1.50 euro per euro that is bet on PSV. At another bookmaker located in Tilburg, the odds are equal: 2 euro for each euro that is bet on either Willem II or PSV. How do you construct an arbitrage? (In the context of bookmaking, a combination of bets that surely wins is known as a "Dutch book".)

c. What is the likely response of the bookmakers in part b. when people start making use of the arbitrage opportunity?

2. A company based in the Eurozone is planning to buy equipment in Japan. Payment will be made in yen, one year from now. To avoid the currency risk involved with the transaction, the company wants to enter a contract in which the rate for conversion from euro to yen in one year's time is already fixed now (a forward exchange contract). Assume

[6] There is an assumption here that market prices are themselves internally consistent, or in other words that market prices do not allow arbitrage. This assumption may not always be satisfied, but normally it should, and we will work on the basis of this assumption.

that the current exchange rate is 160 yen to 1 euro, and that the interest rate in Japan is 1% whereas in the Eurozone it is 5%. Provide an arbitrage argument to determine the one-year forward exchange rate that should hold in the market. In particular, show how an arbitrage opportunity would arise if the forward rate would be either higher or lower than the value that you find.

3. Let c_0 be the price of a European call option and p_0 be the price of a European put option, both on the same asset S, with the same time of maturity T, and the same strike K.[7] Prove that absence of arbitrage implies that

$$c_0 - p_0 = S_0 - d_T K$$

where d_T denotes the discount factor that applies to period T. (The relation above is known as *put-call parity*.)

4. Consider two contracts whose payoffs at time T may depend on the situation at that time. Let the current prices of the contracts be given by C_0^1 and C_0^2, and let their (stochastic) payoffs at time T be denoted by C_T^1 and C_T^2. Assume that it is possible to take positive and negative positions in both contracts.

a. Suppose that the inequality $C_T^2(\omega) \geq C_T^1(\omega)$ holds in all situations ω that may arise at time T.[8] Prove that $C_0^2 \geq C_0^1$ or otherwise there is an arbitrage opportunity.

b. Let C_T be the payoff of a call option with strike K; that is, $C_T = \max(S_T - K, 0)$ where S_T is the value at time T of the underlying asset S. Assume that there is no arbitrage, that interest rates are nonnegative, that the value of the underlying asset is always nonnegative, that $K \geq 0$, and that there are no intermediate cashflows (such as dividends or storage costs) generated by the asset S. Use the result of question a. to prove the following inequalities:

(i) $C_0 \geq 0$

(ii) $C_0 \leq S_0$

(iii) $C_0 \geq S_0 - d_T K$, where d_T is the discount factor that applies to period T.

c. From the above inequalities, derive the inequality $C_0 \geq \max(S_0 - K, 0)$.

Motivation: The final inequality is of interest since it implies that the continuation value of an American call option is at least as large as the value of immediate exercise at time 0.[9] The same reasoning in fact applies at every time $t \leq T$ so that it follows that it is

[7] Call options and put options were introduced in Section 1.2. The value of a call option at the time of maturity T is given by $\max(S_T - K, 0)$; the value of a put option is given by $\max(K - S_T, 0)$.

[8] The notation ω is used here as in probability theory to refer to a possible outcome. There is no relation to initial endowments.

[9] American options differ from their European counterparts in that they may also be exercised before the time at maturity, rather than at the time of maturity.

never advantageous to exercise an American call option early. Consequently, the price of an American call option is the same as the price of its European counterpart. Note that it is assumed here that there are no intermediate cash flows; so if the underlying is stock, this means in particular that it is assumed that the stock does not pay dividends.

5. A "call on a call" is a contract which gives the holder the right to buy at a certain time in the future, for a price that is already fixed now, a call option that expires at some later date. The parameters of the contract are therefore as follows: first time of expiry T_1, second time of expiry T_2, first strike price K_1, second strike price K_2. The option is written on an underlying asset with price S_t at time t. It is assumed that $T_2 > T_1$. At time T_1 the holder has the right to purchase for the price K_1 a European call option on the underlying asset with maturity date T_2 and strike K_2.

Assume that there is a constant continuously compounded interest rate $r \geq 0$ which holds for all maturities, and that there are no dividends or storage costs associated to the underlying asset. Provide an arbitrage argument to show that the price cc_0 at time 0 of the call-on-a-call contract satisfies $cc_0 \geq S_0 - (K_1 + K_2)$.

Chapter 6

Consistent pricing of multiperiod cashflows

6.1 Arbitrage in the bond market

English-Dutch vocabulary for Section 6.1	
annual jaarlijks	**integer** geheel getal
coupon coupon, tussentijdse rentebetaling	**principal** hoofdsom
coupon rate couponrente	**time of maturity** afloopdatum
face value nominale waarde	**time to maturity** resterende looptijd

To study the notion of arbitrage in a mathematical way, both the assets that are available for trading and the admissible trading strategies need to be defined precisely. When these are specified, we say that we have defined a "market", even though in more precise terms, what we define is a mathematical *model* of a market.

In this spirit, let us define a *bond market*. The assets in this market are series of deterministic cashflows which take place at given moments in time. Let us for instance consider only series of cashflows that give rise to payments one year from now, two years from now, and so on up to n years from now where n is a given integer. Such series of deterministic cashflows can be represented by vectors of length n. For instance, if $n = 3$, the vector $(5, 5, 105)$ represents a contract that pays 5 euro one year from now, again 5 euro two years from now, and finally 105 euros three years from now. In financial terminology, such a contract is referred to as a *bond* with a face value of 100 euro, three years to maturity, and a 5% annual coupon. To complete the specification of an asset, also its current price needs to be given. It will be assumed that a finite number (say m) assets are specified by their prices $\pi_1, \ldots, \pi_m$ and their cashflow vectors $f_1, \ldots, f_m$ where each f_i ($i = 1, \ldots, m$) is a row vector of length n.

To complete the specification of the bond market that is considered in this section, we also need to specify what kind of trading strategies are allowed. In general, one can dis-

	Bond 1	Bond 2	Bond 3
Face value	100	100	100
Time to maturity	1 yr.	2 yr.	2 yr.
Coupon	4%	6%	2%
Current price	101	105	97.5

Table 6.1: Available bonds

tinguish between *static* strategies that involve only trading at the current time, and *dynamic* strategies that may involve trading at later times as well. The formulation of dynamic strategies requires a dynamic model for asset prices which gives information not only about current but also about future prices. In this section we only consider static strategies, so that information concerning current prices is sufficient; dynamic trading strategies will be employed in Chapters 8. Furthermore, we assume that it is possible to hold non-integer as well as negative amounts of assets, and that there are no transaction costs; modifications of these assumptions will be discussed in Section 6.5 below. A trading strategy is then just the formation of a portfolio consisting of x_1 units of asset 1,[1] x_2 units of asset 2, and so on up to asset m, where $x_1, x_2, \ldots, x_m$ can be arbitrary real (positive or negative) numbers. The cost of setting up such a portfolio is $x_1\pi_1 + x_1\pi_2 + \cdots + x_m\pi_m$. The portfolio generates the cashflow vector

$$f = x_1 f_1 + x_2 f_2 + \cdots + x_m f_m.$$

It is convenient to use matrix-vector notation, and this will be done systematically below. Specifically we will work with a *cashflow matrix* which is a matrix F consisting of the cashflow vectors of the assets that are available in the market. The cashflow vectors are taken as row vectors so that the cashflow matrix has size $m \times n$. Additionally we have a *price vector*, which contains the prices of the assets and which is a column vector π of length m. Portfolio holdings can be collected in a vector x of length m. The formula above then becomes

$$f = x'F.$$

Note that cashflows have been defined as *row* vectors.

To illustrate the notion of arbitrage in the static bond market context, let us consider a specific example. Suppose that bonds are available in the market as specified in Table 6.1. Given these prices, let us see whether an arbitrage opportunity exists. An arbitrage can be

[1] Since in this chapter we do not consider different agents, we now use subscripts rather than superscripts to refer to the assets in the notation for portfolio holdings.

constructed when there exists a row vector $[x_1\ x_2\ x_3]$ such that

$$\pi_x := [x_1 \ \ x_2 \ \ x_3] \begin{bmatrix} 101 \\ 105 \\ 97.5 \end{bmatrix} \leq 0, \qquad f_x := [x_1 \ \ x_2 \ \ x_3] \begin{bmatrix} 104 & 0 \\ 6 & 106 \\ 2 & 102 \end{bmatrix} \geq 0$$

and moreover the three numbers π_x, $f_x(1)$, and $f_x(2)$ (representing cashflows at times 0, 1, and 2 respectively) are *not* all equal to zero. A *strict* arbitrage can be constructed if there exists a solution x to the inequalities $\pi_x < 0$ and $f_x \geq 0$.

In the above, as usual, the notation $z \geq 0$ when z is a vector means that *all* entries of z are nonnegative. Below we also use the notation $z > 0$, to indicate that *all* entries of the vector z are positive. The notation $z \neq 0$, however, does not mean that all entries of z are nonzero; instead, it means that z is not equal to the zero vector, or in other words, that *at least one* element of z is nonzero.

If we write $x = [x_1\ x_2\ x_3]'$ and define a cashflow matrix F and a price vector π by

$$F = \begin{bmatrix} 104 & 0 \\ 6 & 106 \\ 2 & 102 \end{bmatrix}, \qquad \pi = \begin{bmatrix} 101 \\ 105 \\ 97.5 \end{bmatrix}$$

then the conditions for arbitrage can be written as

$$x'[-\pi \ \ F] \geq 0, \quad x'[-\pi \ \ F] \neq 0. \tag{6.1}$$

The conditions for strict arbitrage are

$$x'\pi < 0, \quad x'F \geq 0. \tag{6.2}$$

These conditions may also be written in the form

$$x'[-\pi \ \ F] \geq 0, \quad x'\pi \neq 0 \tag{6.3}$$

which shows more evidently that the existence of a strict arbitrage is a stronger condition than the existence of an arbitrage (compare (6.3) to (6.1)).

In the case of the example, it is easy to construct an arbitrage and even a strict arbitrage. Note that the matrix $[-\pi \ \ F]$ is invertible:

$$\begin{bmatrix} -101 & 104 & 0 \\ -105 & 6 & 106 \\ -97.5 & 2 & 102 \end{bmatrix}^{-1} = \begin{bmatrix} -0.2857 & 7.5771 & -7.8743 \\ -0.2679 & 7.3586 & -7.6471 \\ -0.2679 & 7.0986 & -7.3671 \end{bmatrix}$$

Consequently, given any vector y of length 3, we can find a vector x of length 3 such that

$$x'[-\pi \;\; F] = y'.$$

In particular this is possible for vectors y that satisfy $y \geq 0$ and $y \neq 0$. Take for instance $y' = [100 \;\; 0 \;\; 0]$. The corresponding vector x is given by

$$x' = y'[-\pi \;\; F]^{-1} = [-28.57 \;\; 757.71 \;\; -787.43].$$

In other words, a strict arbitrage is created by buying 757.71 units of the second bond, and selling 28.57 units of the first bond and 787.43 units of the third bond. This leads to an immediate profit of 100 euro and no net obligations at any time point in the future.[2] Of course this is only one possibility; in fact, *any* series of cashflows at the three time points 0, 1, and 2 can be generated by taking a suitable combination of the three bonds.

The bond market that has been described above is an example of a market in which the assets can be described by means of payoff vectors, and where trading strategies consist of forming a combination of such assets. Another example of such a market is the single-period economy with finitely many states as described in Chapter 3. The available assets in the market can in both cases be summarized by a payoff matrix and an associated price vector, and the strategies can be represented by vectors. In this situation, the general definitions of arbitrage and strict arbitrage given above can be formulated in mathematical terms as follows.

Definition 6.1.1 Consider a market that is given by a payoff matrix $F \in \mathbb{R}^{m \times n}$ and an associated price vector $\pi \in \mathbb{R}^m$. A vector $x \in \mathbb{R}^n$ constitutes an *arbitrage* if

$$x'F \geq 0, \quad x'\pi \leq 0, \quad \text{and} \quad [x'\pi \;\;\; x'F] \neq 0.$$

A vector $x \in \mathbb{R}^n$ represents a *strict arbitrage* if

$$x'F \geq 0 \quad \text{and} \quad x'\pi < 0.$$

A model of a financial market is usually considered implausible if it allows arbitrage. The unlimited profits that are made possible by arbitrage opportunities would create market forces of sufficient size to change prices, and the adjustment of prices would continue until the arbitrage opportunity has disappeared. Therefore, models for markets that have validity for more than a very short period should be arbitrage-free. It is then of interest to have conditions that can guarantee absence of arbitrage. This is the subject of the next section.

6.2 Absence of arbitrage

It seems that it is easier to show that there is a possibility for arbitrage in a given market than to prove that there is no such possibility. After all, the presence of an arbitrage opportunity

[2] The amount of 100 euro is what would be obtained from exact computation. The effect of rounding errors is substantial; the numbers as shown actually lead to a profit of 114 euro.

English-Dutch vocabulary for Section 6.2
discount factor disconteringsfactor; factor waarmee een op een zeker moment in de toekomst plaatsvindende betaling wordt vermenigvuldigd om de huidige waarde te vinden

can be convincingly argued by presenting a row vector x that satisfies the conditions (6.1), or (6.3) in the case of strict arbitrage. To claim that there is no arbitrage opportunity on the basis of the fact that you are not able to find one may be less convincing. However, it turns out that it is possible to argue the *absence* of arbitrage in the same way as the *presence* of arbitrage can be argued, namely by displaying a vector that satisfies certain properties. In the bond market, the vector that will be needed to prove absence of arbitrage can be interpreted as a vector of *discount factors* so that it has a financial meaning in addition to its technical use. The mathematical result that lies at the basis of this all is known as *Stiemke's lemma* and dates from 1915.[3]

Theorem 6.2.1 *Let A be a (not necessarily square) matrix. The matrix equation $x'A = b'$ has no solution for any given vector b such that $b \geq 0$ and $b \neq 0$, if and only if there exists a vector u such that $Au = 0$ and $u > 0$.*

The "if" part of this theorem is easy to prove. Indeed, if there would be vectors x and u such that $x'A \geq 0$, $x'A \neq 0$, $u > 0$, and $Au = 0$, then from the first three of these conditions it would follow that $x'Au > 0$, whereas from the last it follows that $x'Au = 0$; therefore, such vectors x and u cannot exist at the same time. The "only if" part is harder to show. In an appendix to this chapter, a full proof is given of a related result known as *Farkas' lemma*, which will be used below in the context of strict arbitrage; a proof of Stiemke's lemma can be given along similar lines.

To apply Thm. 6.2.1 to the arbitrage condition (6.1), take A equal to $[-\pi \quad F]$. The theorem states that *either* there exists an x such that $x'[-\pi \quad F] = y'$ for some nonzero $y \geq 0$ (i.e. there is an arbitrage opportunity), *or* the equation $[-\pi \ \ F]u = 0$ has a solution $u > 0$. In other words, the market does *not* allow arbitrage opportunities if and only if there is a vector u that satisfies the conditions

$$[-\pi \ \ F]u = 0, \qquad u > 0. \tag{6.4}$$

If we cannot find such a vector, we should be able to find a vector x that satisfies (6.1). Therefore both the presence and the absence of arbitrage can be proved by displaying a vector that satisfies certain requirements. As will be discussed below, there is a similar statement that applies to the absence or presence of *strict* arbitrage.

The bond market interpretation of the vector u in (6.4) is that it provides discount factors. To be more precise, note first of all that if u satisfies (6.4), then the same is true for λu

[3] Erich Stiemke (1892–1915), German mathematician. Stiemke died in military service in World War I.

where λ can be any positive number. Therefore we may assume without loss of generality that the first entry of u is equal to 1. This is the entry that corresponds to the first column of the matrix $[-\pi \quad F]$, namely the price vector. The remaining entries correspond to the columns of the cashflow matrix F, which in turn relate to the different time points at which cashflows take place. Let us write $u = [1 \quad d_1 \quad \cdots \quad d_n]'$, and $[d_1 \ \cdots \ d_n]' = d$. The equality $[-\pi \ \ F]u = 0$ then becomes

$$\pi = Fd. \tag{6.5}$$

In other words, the market defined by the price vector π and the corresponding cashflow matrix F is free of arbitrage if there exists a vector $d = [d_1 \ \cdots \ d_n]'$ with positive entries such that for every product i ($i = 1, \ldots, m$) with cashflows $[f_{i1}, \ldots, f_{in}]$ the following relation holds:

$$\pi_i = \sum_{j=1}^{n} f_{ij} d_j. \tag{6.6}$$

In particular, if one of the given products is a zero-coupon bond with face value 1, maturing at the time T_j that corresponds to the j-th column of the cashflow matrix, then the price of that bond is given by d_j. In other words, the entries of the vector d can be interpreted as the *discount factors* that follow from the prices of the products defined by the cashflows given by the rows of the cashflow matrix F. Given this interpretation, equation (6.6) is nothing else but the *net present value formula*. The standard use of this formula is to obtain prices when discount factors are given; here the prices are given, and the discount factors are obtained from them. If the equation (6.5) for given π has a unique solution d, we may therefore call d the vector of discount factors that are *implied* by the prices in the vector π.

The conclusion from the application of Thm. 6.2.1 can be formulated as follows.

Fact 6.2.2 *In a market defined by a cashflow matrix F and a corresponding price vector π, there is absence of arbitrage if and only if there exists a vector d of positive discount factors such that $\pi = Fd$.*

Discount factors are usually expressed in terms of *interest rates*. The *continuously compounded interest rate* r_T^{c} for maturity T is defined as the solution of the equation

$$e^{-r_T^{\mathrm{c}} T} = d_T.$$

In other words, $r_T^{\mathrm{c}} = -\frac{1}{T} \ln d_T$. The *discretely compounded interest rate* r_T^{d} for maturity T is defined as the solution of the equation

$$\frac{1}{(1 + r_T^{\mathrm{d}})^T} = d_T.$$

In other words, $r_T^{\mathrm{d}} = d_T^{-1/T} - 1$. The positivity of discount factors guarantees that these interest rates can indeed be defined in this way. It does *not* guarantee that interest rates are positive. As is seen from the formulas above, the interest rate (either continuously or discretely compounded) is positive if and only if the discount factor is less than one. We return to the issue of nonnegativity of interest rates in Section 6.5 below.

Now let us consider the notion of *strict* arbitrage. As already mentioned above, in this case we can use *Farkas' lemma*[4]. This theorem is actually older (published in 1902) than Stiemke's lemma and it is more famous, because of many applications for instance in optimization.

Theorem 6.2.3 *Let a matrix A of size $m \times n$ and a vector b of length m be given. There exists no vector x such that $x'A \geq 0$ and $x'b < 0$ if and only if there exists a vector $y \geq 0$ such that $Ay = b$.*

The alternatives expressed by this theorem are:

- *either* there exists a vector x of length m such that $x'A \geq 0$ and $x'b < 0$,
- *or* there exists a vector y of length n such that $y \geq 0$ and $Ay = b$.

As in the case of Stiemke's lemma, it is easy to see that the two alternatives cannot be true at the same time. If there would exist vectors x and y such that both conditions are fulfilled, then on the one hand $x'Ay = x'b < 0$ on the basis of $Ay = b$ and $x'b < 0$, but on the other hand $x'Ay \geq 0$ because $x'A \geq 0$ and $y \geq 0$. The difficult part is to show that at least one of the alternatives must be fulfilled. A proof of Farkas' lemma, with a financial interpretation of every step in the proof, is provided in an appendix (Section 6.7).

To apply the Farkas lemma to conditions for absence of strict arbitrage, let b be equal to the price vector π, and let the matrix A be given by the cashflow matrix F. Then alternative (i) corresponds to the case in which there is a strict arbitrage. This case is excluded by alternative (ii), which (writing d instead of y now) can be written as $\pi = Fd$. Therefore, the conclusion is:

Fact 6.2.4 *In a market defined by a cashflow matrix F and a corresponding price vector π, there is absence of strict arbitrage if and only if there exists a vector d of nonnegative discount factors such that $\pi = Fd$.*

Because strict arbitrage is a stronger notion than arbitrage, the condition for *absence* of strict arbitrage must be *weaker* than the condition for absence of arbitrage. Indeed, the condition in Fact 6.2.4 is weaker than the condition in Fact 6.2.2. If there is no positive vector d that satisfies the equation $\pi = Fd$, but there is a nonnegative solution (with some zero entries), then the market admits an arbitrage opportunity but not a strict arbitrage opportunity.

[4] Gyula Farkas (1847–1930), Hungarian mathematician and physicist.

Example 6.2.5 In the previous section, we found that the market defined by

$$\pi = \begin{bmatrix} 101 \\ 105 \\ 97.5 \end{bmatrix}, \qquad F = \begin{bmatrix} 104 & 0 \\ 6 & 106 \\ 2 & 102 \end{bmatrix}$$

allows arbitrage. Suppose now that we want to re-price the two-year bond with 2% coupon rate in such a way that the arbitrage opportunity disappears. To see whether this is possible, we first of all need to check that the market consisting of only the first two contracts is free of arbitrage. Therefore, consider

$$\pi_1 = \begin{bmatrix} 101 \\ 105 \end{bmatrix}, \qquad F_1 = \begin{bmatrix} 104 & 0 \\ 6 & 106 \end{bmatrix}$$

and write down the equation

$$\begin{bmatrix} 104 & 0 \\ 6 & 106 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 101 \\ 105 \end{bmatrix}.$$

There is one solution, which is given by $d_1 = 0.9712$, $d_2 = 0.9356$. These numbers are positive and so it follows from Fact 6.2.2 that the market given by the cashflow matrix F_1 and the corresponding price vector π_1 is free of arbitrage. The price of the third contract must be determined by the same discount factors d_1 and d_2, or otherwise the equation $\pi = Fd$ has no solutions and consequently an arbitrage would exist. Therefore, the market-consistent price of the cashflow $[2 \;\; 102]$ is given by

$$0.9712 \cdot 2 + 0.9356 \cdot 102 = 97.3730.$$

This is the only price that prevents arbitrage. In particular, the price 97.5 that we considered before gives rise to an arbitrage opportunity, as we have seen.

From the discount factors, we can obtain the corresponding interest rates. Suppose that we work with discretely compounded rates; then

$$\frac{1}{1 + r_1} = d_1, \qquad \frac{1}{(1 + r_2)^2} = d_2.$$

Given the above values of the discount factors d_1 and d_2, we find from these relations that the one-year interest rate r_1 is 2.97%, and the two-year interest rate r_2 is 3.38%. Note that the principle of absence of arbitrage only implies that for each maturity there must be a single interest rate; interest rates for different maturities can be different, as is seen in the example and as is also usually observed in practice.

English-Dutch vocabulary for Section 6.3

complete market volledige markt
replicable repliceerbaar
replicate repliceren, nabouwen

6.3 The replication theorem

Given a collection of series of future cashflows $(f_{11}, \dots, f_{1n}), \dots, (f_{m1}, \dots, f_{mn})$, we say that a series of cashflows $\hat{f} = (\hat{f}_1, \dots, \hat{f}_n)$ is *replicable* if there exist numbers $x_1, \dots, x_m$ such that

$$\sum_{j=1}^{m} x_j f_{ji} = \hat{f}_i \quad \text{for all } i = 1, \dots, n.$$

In terms of linear algebra, this just means that the row vector $(\hat{f}_1, \dots, \hat{f}_n)$ is a linear combination of the row vectors $(f_{11}, \dots, f_{1n})$, $\dots$, $(f_{m1}, \dots, f_{mn})$. In economic terms, it means that we can form a portfolio of the contracts that generate the series of cashflows $(f_{11}, \dots, f_{1n}), \dots, (f_{m1}, \dots, f_{mn})$, in such a way that this portfolio generates exactly the cashflows $(\hat{f}_1, \dots, \hat{f}_n)$; in other words, the contract that generates these cashflows is *replicated* by the portfolio.

Replication has consequences for pricing. Assume that prices are given of a number of contracts that generate deterministic cashflows; let π (a vector of length m) denote the price vector, and let F, a matrix of size $m \times n$, represent the cashflow matrix. Also assume that the given contracts are independent, that is, none of the contracts can be replicated by a combination of the other contracts, and that all contracts can be held in arbitrary (positive or negative) quantities. Consider now a contract that generates a series of cashflows $\hat{f} = (\hat{f}_1, \dots, \hat{f}_n)$. If the row vector $\hat{f}$ is replicable by means of the contracts with given prices, so that $\hat{f} = x'F$ for some m-vector x, then the price of the contract that generates the series of cashflows $\hat{f}$ must be equal to $x'\pi$, or otherwise there is an arbitrage opportunity. This suggests a connection between uniqueness of discount factors and replicability of all assets, as is contained in the following statement.

Fact 6.3.1 *The following conditions are equivalent for a cashflow matrix F of size $m \times n$:*

(i) discount factors are unique in the sense that for any given price vector $\pi \in \mathbb{R}^m$, there exists at most one vector $d \in \mathbb{R}^n$ such that $Fd = \pi$;

(ii) all series of cashflows can be replicated; in other words, for each row vector $\hat{f}$ of length n there exists a vector x of length m such that $\hat{f} = x'F$.

This is called the *replication theorem*. Note that in condition (i), the possibility is left open that there exists *no* vector d such that $Fd = \pi$. This happens when the rows of the matrix F are dependent, but the price vector π doesn't reflect that — for instance, when the second row of F is twice the first row, but π_2 is not equal to $2\pi_1$. In such a situation there is an arbitrage opportunity. Indeed, the condition of Fact 6.2.2 is not satisfied.

Proof To prove the replication theorem, we show that the two conditions in the theorem are both equivalent to

(iii) the matrix F has rank n.

This is seen as follows:

for any π such that a solution of $Fd = \pi$ exists, the solution is unique

$\Leftrightarrow$ the columns of the matrix F are independent vectors

$\Leftrightarrow$ the rank of the matrix F is n

$\Leftrightarrow$ the rows of F span the n-dimensional space

$\Leftrightarrow$ every series of cashflows can be replicated.

□

A market that satisfies the conditions of the replication theorem is called a *complete market*. Note that a necessary condition for completeness is that $m \geq n$; in other words, the number of rows of the matrix F, which is equal the number of assets in the market, should be at least equal to the number of columns of F, which is the same as the number of payment dates.

A property that is in some sense a mirror image of completeness is the property of *uniqueness of replication*. A market specified by a cashflow matrix F and a price vector π is said to satisfy this property if, whenever a cashflow can be replicated by means of the assets specified in the matrix F, the replication can be done in only one way. In more mathematical terms, this means: for each row vector $\hat{f}$ of length n, there exists at most one vector of holdings $x \in \mathbb{R}^m$ such that $x'F = \hat{f}$. Equivalently, the rows of the matrix F are independent, or in still other terms, the rank of the $m \times n$ matrix F is equal to m. Just as in the case of completeness, uniqueness of replication in fact depends only on the cashflow matrix F, not on the price vector π.

The *law of one price* states that financial products with identical cashflows in all possible states of the world should have identical prices. In a market characterized by a cashflow matrix F and a price vector π, this property means that the equality $x_1'\pi = x_2'\pi$ should hold whenever x_1 and x_2 are vectors of holdings such that $x_1'F = x_2'F$. An equivalent statement is that the relation $x'F = 0$ should imply $x'\pi = 0$. This property holds if and only if there exists a solution d (not necessarily with positive or nonnegative entries) of the vector-matrix equation $Fd = \pi$. Indeed, if there exists a solution to the equation $Fd = \pi$, then $x'F = 0$ implies $x\pi = xFd = 0$. Conversely, suppose that the equation $Fd = \pi$ does not have a solution. This means that the column vector π cannot be written as a linear combination of the columns of the matrix F, or in other words that π does not belong to the column space

Market given by $F \in \mathbb{R}^{m \times n},\ \pi \in \mathbb{R}^m$ Equation $\pi = Fd$			compl.	uniq. repl.	abs. of arb.	abs. of str. arb.	L1P
$n = \operatorname{rk} F$	$m = \operatorname{rk} F$	$d > 0$	yes	yes	yes	yes	yes
		$d \not> 0,\ d \geq 0$	yes	yes	no	yes	yes
		$d \not\geq 0$	yes	yes	no	no	yes
	$m > \operatorname{rk} F$	$d > 0$	yes	no	yes	yes	yes
		$d \not> 0,\ d \geq 0$	yes	no	no	yes	yes
		$d \not\geq 0$	yes	no	no	no	yes
		$\nexists\, d$	yes	no	no	no	no
$n > \operatorname{rk} F$	$m = \operatorname{rk} F$	$\exists\, d > 0$	no	yes	yes	yes	yes
		$\nexists\, d > 0,\ \exists\, d \geq 0$	no	yes	no	yes	yes
		$\nexists\, d \geq 0$	no	yes	no	no	yes
	$m > \operatorname{rk} F$	$\exists\, d > 0$	no	no	yes	yes	yes
		$\nexists\, d > 0,\ \exists\, d \geq 0$	no	no	no	yes	yes
		$\nexists\, d \geq 0,\ \exists\, d$	no	no	no	no	yes
		$\nexists\, d$	no	no	no	no	no

Table 6.2: Market types

of F. Say that the dimension of the column space of F is m'; we must have $m' < m$, because otherwise there can be no m-vector π outside the column space of F. Choose a set of independent columns of F, and let these columns be denoted by $F_1, \ldots, F_{m'}$. We can form a basis of the m-dimensional space $\mathbb{R}^m$ by taking first the vectors $F_1, \ldots, F_{m'}$, then the vector π, and possibly other vectors if needed (in case $m' < m - 1$) to complete the basis. These basis vectors together form an $m \times m$ matrix, say A, which is invertible by construction. Consider the $(m' + 1)$-th row of the inverse matrix A^{-1}, and write this row as the transpose of a column vector x. Then we have $x'\pi = 1$ by definition of the inverse. Moreover, the definition of the inverse also implies that $x'F_j = 0$ for $j = 1, \ldots, m'$. Since the vectors F_j form a basis for the column space of F, it follows that $x'F = 0$. In other words, we have found a vector x such that $x'F = 0$ and $x'\pi \neq 0$.

We have now identified five properties that may or may not be satisfied by markets that are characterized by a cashflow matrix F and a price vector π: completeness, uniqueness of replication, absence of arbitrage, absence of strict arbitrage, and validity of the law of one

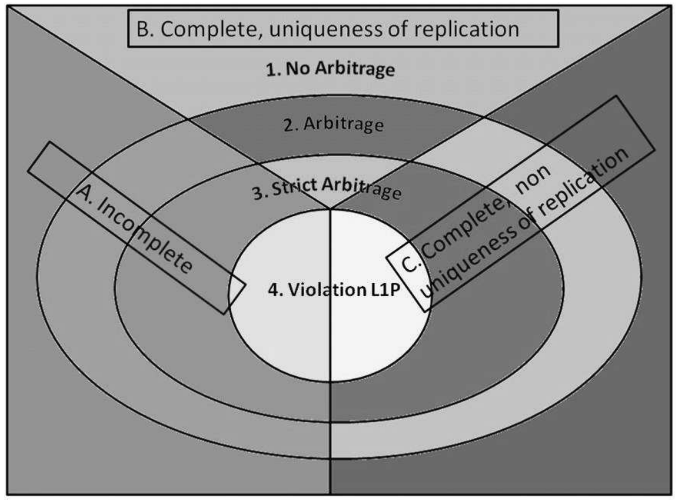

Figure 6.1: Market types

price. A summary of the conditions under which these properties hold is given in Table 6.2. There are fourteen different cases to be distinguished. A representation in more graphical form in Fig. 6.1.[5] In the figure, incomplete markets have not been subdivided according to the property of uniqueness of replication.

6.4 Incomplete markets

In Example 6.2.5, the discount factors corresponding to cashflow times were determined uniquely on the basis of the prices of two contracts, namely the one-year bond and the two-year bond with 6% coupon rate. This was a consequence of the fact that the prices of two independent products were given and we were looking for discount factors corresponding to two time points, namely one year and two years from now. If there are more times at which cashflows take place than there are products whose prices are given, then it is not possible to determine the discount factors exactly, and no exact price can be determined for a new product (i.e. one that is not a linear combination of the products whose prices are given). In this case we speak of an *incomplete market*: not all products can be priced on the basis of available market data.

Although in an incomplete market prices in general cannot be determined exactly, arguments based on absence of arbitrage can still be used to determine *lower bounds* and *upper bounds* for prices. To see how this works, let $T_1, \dots, T_n$ be given times at which cashflows will take place, and suppose that prices are given of m contracts. The contracts produce deterministic cashflows that occur at the times $T_1, \dots, T_n$. The cashflows of all contracts at

[5] Thanks to dr. Erwin Charlier for providing this figure.

all times are specified in the cashflow matrix F which is of size $m \times n$. The corresponding price vector π is a vector of length m. Suppose that a contract is presented whose associated series of cashflows is not a linear combination of the rows of the matrix F. Denote the series of cashflows of this contract by $\hat{f}$; this is a row vector of length n.

We assume that the market given by the cashflow matrix F and the corresponding price vector π is free of arbitrage. If a new contract $\hat{f}$ is introduced with price $\hat{\pi}$, it might be that an arbitrage opportunity is created; in such a case one may speak of a mispricing. This may happen in two ways. If the price of the new contract is too low, then it may be possible to create an arbitrage involving *positive* ("*long*") position in $\hat{f}$, since in such a case it could be attractive to buy the contract. On the other hand, if the price is too high, this may create a possibility for an arbitrage based on selling the contract, in other words, taking a *negative* ("*short*") position in $\hat{f}$. To formulate this in mathematical terms, write f_j for the j-th row of the matrix F (i.e. the series of cashflows generated by the j-th contract). The condition for absence of strict arbitrage based on a long position in the new contract $\hat{f}$ can then be stated as follows: the implication

$$\hat{f} + \sum_{j=1}^{m} x_j f_j \geq 0 \quad \Longrightarrow \quad \hat{\pi} + \sum_{j=1}^{m} x_j \pi_j \geq 0 \tag{6.7}$$

should hold for all $x_1, \ldots, x_m$. The portfolio weight of $\hat{f}$ has here been set equal to 1 without loss of generality; indeed, the weight should be positive because the portfolio is based on a long position in $\hat{f}$, and an arbitrage portfolio multiplied by any positive constant is still an arbitrage portfolio. Likewise, the condition for absence of strict arbitrage based on a short position is

$$-\hat{f} + \sum_{j=1}^{m} x_j f_j \geq 0 \quad \Longrightarrow \quad -\hat{\pi} + \sum_{j=1}^{m} x_j \pi_j \geq 0 \tag{6.8}$$

for all $x_1, \ldots, x_m$. To write the conditions for absence of strict arbitrage as price constraints, define the following numbers:

$$\underline{\pi}(\hat{f}) = \max_{x} \left\{ x'\pi \mid x'F \leq \hat{f} \right\} \tag{6.9}$$

$$\overline{\pi}(\hat{f}) = \min_{x} \left\{ x'\pi \mid x'F \geq \hat{f} \right\}. \tag{6.10}$$

The conditions (6.7) and (6.8) can then be rewritten in a compact form as

$$\underline{\pi}(\hat{f}) \leq \hat{\pi} \leq \overline{\pi}(\hat{f}). \tag{6.11}$$

The quantities $\overline{\pi}(\hat{f})$ and $\underline{\pi}(\hat{f})$ are called the *upper arbitrage bound* and the *lower arbitrage bound* respectively. To compute these quantities is a *linear programming problem*. It follows from the *duality theorem* of linear programming that these two numbers can also be obtained as follows:

$$\underline{\pi}(\hat{f}) = \min\{\hat{f}d \mid Fd = \pi,\ d \geq 0\} \tag{6.12}$$

$$\overline{\pi}(\hat{f}) = \max\{\hat{f}d \mid Fd = \pi,\ d \geq 0\}. \tag{6.13}$$

This means that the lower bound $\underline{\pi}(\hat{f})$ and the upper bound $\overline{\pi}(\hat{f})$ can be computed by finding the set of all solutions of the vector-matrix equation $Fd = \pi$ that satisfy the nonnegativity constraint $d \geq 0$. This may be easier than to solve the primal optimization problems (6.9) and (6.10), especially when the market is nearly complete in the sense that the difference between the dimensions of all the space of all replicable assets and the space of all assets is small.

The following fact shows that it is always possible to find a price for the new contract that prevents strict arbitrage, if the original market is arbitrage-free.

Fact 6.4.1 *If the market defined by the cashflows $f_1, \ldots, f_m$ and the corresponding prices $\pi_1, \ldots, \pi_m$ is free of arbitrage, then $\underline{\pi}(\hat{f}) \leq \overline{\pi}(\hat{f})$ for all $\hat{f}$.*

Proof For a proof by contradiction, suppose that $\underline{\pi}(\hat{f}) > \overline{\pi}(\hat{f})$, i.e.

$$\max\{-\textstyle\sum_{i=1}^m x_j\pi_j \mid \sum_{j=1}^m x_j f_j \geq -\hat{f}\} > \min\{\sum_{i=1}^m x_j\pi_j \mid \sum_{j=1}^m x_j f_j \geq \hat{f}\}.$$

Then there exist $x_1, \ldots, x_m$ and $\tilde{x}_1, \ldots, \tilde{x}_m$ such that

$$-\sum_{i=1}^m x_j\pi_j > \sum_{i=1}^m \tilde{x}_j\pi_j, \quad \sum_{j=1}^m x_j f_j \geq -\hat{f}, \quad \sum_{j=1}^m \tilde{x}_j f_j \geq \hat{f}.$$

But then

$$\sum_{j=1}^m (x_j + \tilde{x}_j)\pi_j < 0 \text{ and } \sum_{i=1}^m (x_j + \tilde{x}_j) f_j \geq 0$$

which contradicts the assumption that the market defined by the f_j's is free of arbitrage. □

The statement above provides assurance that the pricing interval $[\underline{\pi}(\hat{f}), \overline{\pi}(\hat{f})]$ is indeed an interval. The length of the interval can be zero, namely when the payoff $\hat{f}$ is replicable. In a sense the width of the interval $[\underline{\pi}(\hat{f}),\, \overline{\pi}(\hat{f})]$ is an indication of the degree of non-replicability of a given payoff. When the interval is wide, little pricing information is obtained from the arbitrage bounds. In such cases it may be possible to bring in additional considerations which help to reduce the length of the interval of reasonable prices.

Example 6.4.2 Suppose that prices are given of a two-year bond with 4% coupon rate and of a three-year bond with 5% coupon rate, both with face value 100, as follows:

$$\pi = \begin{bmatrix} 100 \\ 102 \end{bmatrix}, \quad F = \begin{bmatrix} 4 & 104 & 0 \\ 5 & 5 & 105 \end{bmatrix}.$$

What can we say about the price of a zero-coupon three-year bond with face value 100? To determine the arbitrage bounds, we can solve the linear programming problems contained in (6.9) and (6.10). Since the market would be complete if we would know the price of the three-year zero coupon bond, a specific approach that is convenient in this case is to compute the implied discount rates as functions of the unknown price. The range of values of the price of the bond for which the implied discount factors are positive is the range of prices that exclude arbitrage. Let the price of the zero-coupon bond be denoted by x. Then the discount factors for one, two, and three years respectively are determined by the equations

$$\begin{aligned} 4d_1 + 104d_2 &= 100 \\ 5d_1 + 5d_2 + 105d_3 &= 102 \\ 100d_3 &= x. \end{aligned}$$

After some manipulation, we find:

$$\begin{aligned} d_1 &= 0.208(102 - 1.05x) - 1 \\ d_2 &= -0.008(102 - 1.05x) + 1 \\ d_3 &= 0.01x. \end{aligned}$$

All three discount factors should be positive, which means that all of the following three inequalities must be satisfied:

$$\begin{aligned} &x < \frac{102 - 1/0.208}{1.05} = 92.5641 \\ &x > \frac{102 - 1/0.008}{1.05} = -21.9048 \\ &x > 0. \end{aligned}$$

Therefore, the arbitrage-free interval for the price of the zero-coupon bond is $(0, 92.5641)$. Apparently even very low (but still positive) prices of the zero-coupon bond do not lead to arbitrage opportunities.

Requiring that interest rates must be nonnegative means that the discount factors d_1, d_2 and d_3 cannot exceed 1. This inequality constraint can be derived from an arbitrage argument, but an extension of the framework used so far is needed; this will be worked out in the next section. If we impose that $d_i \leq 1$ for $i = 1, 2, 3$, then the following three inequalities for the bond price x are obtained:

$$\begin{aligned} &x \geq \frac{102 - 2/0.208}{1.05} = 87.9853 \\ &x \leq \frac{102}{1.05} = 97.1429 \\ &x \leq 100. \end{aligned}$$

With the added constraint, the arbitrage-free interval shrinks to $[87.9853, 92.5641)$. The implied interest rates corresponding to various prices of the zero-coupon bond are plotted in

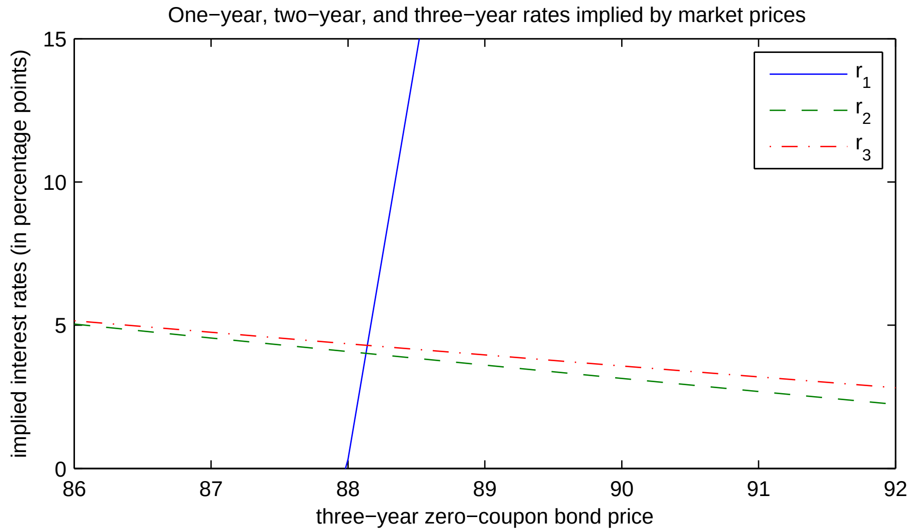

Figure 6.2: Implied interest rates

Fig. 6.2. Especially the implied one-year interest rate is very sensitive to the price, whereas the implied two-year and three-year rates vary more gradually. To get a term structure that is not too unlikely (meaning that interest rates for maturities one year apart should not differ by more than, say, 2 percentage points), the price of the three-year zero-coupon bond should be equal to 88.16, plus or minus eight cents at most.

If we look at arbitrage instead of strict arbitrage, then usually we need to replace non-strict inequalities by strict inequalities. This is seen also in the case of incomplete markets.

Fact 6.4.3 *Suppose that the market defined by the cashflows $f_1, \dots, f_m$ and the corresponding prices $\pi_1, \dots \pi_m$ is free of arbitrage, and suppose that $\hat{f}$ is linearly independent of $f_1, \dots, f_m$. The market defined by the cashflows $f_1, \dots, f_m$ and $\hat{f}$ is then free of arbitrage if and only if*

$$\underline{\pi}(\hat{f}) < \hat{\pi} < \overline{\pi}(\hat{f}). \tag{6.14}$$

Proof We first show that there can be no arbitrage if (6.14) holds. The proof proceeds by contradiction. Suppose there would be an arbitrage opportunity in the extended market. By definition, this means that there exist $\hat{x}$ and x such that

$$\hat{x}\hat{\pi} + x'\pi \leq 0, \quad \hat{x}\hat{f} + x'F \geq 0 \tag{6.15}$$

and not both $\hat{x}\hat{\pi} + x'\pi = 0$ and $\hat{x}\hat{f} + x'F = 0$. Note that $\hat{x}$ must be nonzero, because otherwise there would be arbitrage in the original market with cashflows $f_1, \dots, f_m$. Suppose for instance that $\hat{x} > 0$; then we can without loss of generality take $\hat{x} = 1$. We then have $x'F \geq -\hat{f}$ by the second condition in (6.15). By the definition of $\underline{\pi}(\hat{f})$ in (6.9) and the assumption that (6.14) holds, it follows that $\hat{\pi} > -x'\pi$. But this contradicts the first condition in (6.15). Likewise, the assumption $\hat{x} < 0$ leads to a contradiction. This concludes the first part of the proof.

The second part of the proof consists of showing that there is arbitrage if the condition (6.14) does not hold. Suppose for instance that $\hat{\pi} \geq \overline{\pi}(\hat{f})$ (the proof for the case in which $\hat{\pi} \leq \underline{\pi}(\hat{f})$ is similar). Let x achieve the minimum in (6.10), so that $\overline{\pi}(\hat{f}) = x'\pi$. We then have $-\hat{\pi} + x'\pi \leq 0$ and $-\hat{f} + x'F \geq 0$. Moreover, since it was assumed that $\hat{f}$ is linearly independent of $f_1, \dots, f_m$, we must have $-\hat{f} + x'F \neq 0$ so that indeed an arbitrage has been constructed. □

In the case in which the new cashflow $\hat{f}$ is independent of the existing cashflows $f_1, \dots, f_m$, the conditions for absence of arbitrage are consequently the same as the conditions for absence of strict arbitrage, except that the non-strict inequalities for absence of strict arbitrage are replaced by strict inequalities for absence of arbitrage. When the new cashflow $\hat{f}$ does depend linearly on $f_1, \dots, f_m$ however, the condition for absence of arbitrage is identical to the condition of absence of strict arbitrage: we should have $\hat{\pi} = x'\pi$ where x is such that $\hat{f} = x'F$. In this case we have $\underline{\pi}(\hat{f}) = \overline{\pi}(\hat{f}) = x'\pi$.

To show that it is always possible to find a price for the new cashflow $\hat{f}$ that does not introduce an arbitrage into the market, we still need to show that it cannot happen that $\underline{\pi}(\hat{f}) = \overline{\pi}(\hat{f})$ while $\hat{f}$ is linearly independent of $f_1, \dots, f_m$. The proof of this is given below.

Fact 6.4.4 *Suppose that the market defined by the cashflows $f_1, \dots, f_m$ and the corresponding prices $\pi_1, \dots \pi_m$ is free of arbitrage. If for some cashflow $\hat{f}$ we have $\underline{\pi}(\hat{f}) = \overline{\pi}(\hat{f})$, then $\hat{f}$ can be written as a linear combination of $f_1, \dots, f_m$.*

Proof Under the stated condition, there exist $x_1, \dots, x_m$ and $\tilde{x}_1, \dots, \tilde{x}_m$ such that

$$-\sum_{i=1}^{m} x_j \pi_j = \sum_{i=1}^{m} \tilde{x}_j \pi_j, \quad \sum_{j=1}^{m} x_j f_j \geq -\hat{f}, \quad \sum_{j=1}^{m} \tilde{x}_j f_j \geq \hat{f}.$$

It follows that

$$\sum_{j=1}^{m} (x_j + \tilde{x}_j)\pi_j = 0 \text{ and } \sum_{i=1}^{m} (x_j + \tilde{x}_j) f_j \geq 0.$$

Since it was assumed that the market given by the cashflow matrix F is free of arbitrage, this implies that $(x + \tilde{x})'F = 0$. We can now write

$$\hat{f} \leq \tilde{x}'F = -x'F \leq \hat{f}.$$

Both inequalities must therefore be equalities, and in particular we have $\hat{f} = -x'F$. □

6.5 Inequality constraints

So far it has been assumed that one can take both negative and positive positions in all contracts. This is a liquidity assumption. In this section we take a small excursion into the

English-Dutch vocabulary for Section 6.5
term deposit termijndeposito

modeling of markets that are not fully liquid. We build a model that allows for contracts that can only be held in nonnegative quantities. The following example shows how the extended freedom of modeling can be used, for instance, to describe a situation in which the rate for borrowing is different from the rate for lending.

Example 6.5.1 Suppose that the borrowing rate for one-year maturity is 3.2% whereas the lending rate is 3.1%. This means that you pay 3.2% interest rate if you borrow money, and you get 3.1% if you lend it to someone else (for instance put it into a term deposit). This can be expressed as follows: the cashflow 1031 euro taking place one year from now is available for a price 1000 euro to be paid now, and the cashflow -1032 is available for the price -1000 euro (which means that you can get 1000 euro now by promising to pay 1032 euro one year from now). Both cashflows are only available in nonnegative quantities. Under this restriction, it is not possible to create an arbitrage.

Models in which some assets can only be held in nonnegative quantities are also useful to describe *inequalities* that can be derived from the principle of absence of arbitrage. A standard example of this in the bond market is the rule which states that interest rates must be nonnegative. As already mentioned above, the nonnegativity of interest rates cannot be derived from the no-arbitrage conditions in Fact 6.2.2. However, it is possible to show that interest rates must be nonnegative by introducing an asset that can only be held in nonnegative quantities, namely cash; see Example 6.5.5 below.

To get conditions for absence of arbitrage when some contracts can only be held in nonnegative quantities, we can again a theorem of the alternative. The following extension of Stiemke's lemma is a special case of what is known as Tucker's theorem of the alternative,[6] which dates from 1956.

Theorem 6.5.2 *Let A and B be given matrices with the same number of rows. There does not exist a vector x such that $x'A = b'$ and $x'B = c'$ for any vectors b and c such that $b \geq 0$, $c \geq 0$, and $b \neq 0$, if and only if there exist vectors $y > 0$ and $u \geq 0$ such that $Ay + Bu = 0$.*

To express this more clearly as a theorem of the alternative, the statement of the theorem can be reformulated as follows:

(i) *either* there exists a vector x such that $x'A \geq 0$, $x'B \geq 0$, and $x'A \neq 0$,

(ii) *or* there exist vectors $y > 0$ and $u \geq 0$ such that $Ay + Bu = 0$.

[6] Albert W. Tucker (1905–1995), Canadian/US mathematician.

The theorem is applied to conditions for absence of arbitrage in the following way. Let F_1 be the cashflow matrix corresponding to the contracts that are available in arbitrary (positive or negative) quantities, and let F_2 be the cashflow matrix for the contracts that are available only in nonnegative quantities. The matrix F_1 is of size $m_1 \times n$, and F_2 is of size $m_2 \times n$; the column indices $1, \ldots, n$ of both matrices refer to the times $T_1, \ldots, T_n$ at which cashflows take place. Let π_1 and π_2 denote the corresponding price vectors of length m_1 and m_2 respectively. An arbitrage opportunity exists when there exist vectors $x_1 \in \mathbb{R}^{m_1}$ and $x_2 \in \mathbb{R}^{m_2}$, with $x_2 \geq 0$, such that

$$x_1'\pi_1 + x_2'\pi_2 \leq 0, \qquad x_1'F_1 + x_2'F_2 \geq 0$$

and the left hand sides are not both equal to zero. Now, define

$$A = \begin{bmatrix} -\pi_1 & F_1 \\ -\pi_2 & F_2 \end{bmatrix}, \qquad B = \begin{bmatrix} 0_{m_1 \times m_2} \\ I_{m_2 \times m_2} \end{bmatrix}.$$

The matrix A is of size $(m_1+m_2)\times n$, and the matrix B has size $(m_1+m_2)\times m_2$. With these definitions, alternative (i) above gives exactly the conditions for arbitrage. Consequently, alternative (ii) gives the conditions for *absence* of arbitrage. Note that, if y and u satisfy the requirements of alternative (ii), then the same holds for λy and λu where λ is a positive constant. Therefore we can assume without loss of generality that the vector y is of the form $y' = [1 \;\; d_1 \;\; \ldots \;\; d_n]$ with $d_i > 0$ for all $i = 1, \ldots, n$. Consequently, we find that absence of arbitrage holds when there is a positive vector d and a nonnegative vector u such that

$$\pi_1 = F_1 d, \qquad \pi_2 = F_2 d + u.$$

To say that $\pi_2 = F_2 d + u$ for some nonnegative vector u is the same as to say that $\pi_2 \geq F_2 d$. Therefore the conclusion can be stated as follows.

Fact 6.5.3 *Let a market be defined by cashflow matrices F_1 and F_2 and by corresponding price vectors π_1 and π_2, where contracts of the first kind can be held in arbitrary quantities (positive or negative), and contracts of the second kind can only be held in nonnegative quantities. Absence of arbitrage holds in this market if and only if there exists a vector d of strictly positive discount factors such that*

$$\pi_1 = F_1 d \quad \text{and} \quad \pi_2 \geq F_2 d. \tag{6.16}$$

We can also formulate conditions for absence of strict arbitrage. In fact for this purpose we can use Farkas' lemma directly. Define

$$A = \begin{bmatrix} F_1 & 0_{m_1 \times m_2} \\ F_2 & I_{m_2 \times m_2} \end{bmatrix}, \qquad b = \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix}.$$

An application of Thm. 6.2.3 leads to the following result.

Fact 6.5.4 *In the same situation as in Fact 6.5.3, absence of strict arbitrage holds if and only if there exists a vector* d *of nonnegative discount factors such that*

$$\pi_1 = F_1 d \quad \text{and} \quad \pi_2 \geq F_2 d. \tag{6.17}$$

Example 6.5.5 Suppose that a cashflow of 100 euro at time T is available now for the price of 100 euro. A contract like that is simply obtained by holding cash (in a safe place). The availability in nonnegative quantities of such an asset implies, by Fact 6.5.3, that to avoid arbitrage the discount rate d_T that is used for maturity T must satisfy

$$100 \geq 100 d_T$$

or in other words

$$d_T \leq 1. \tag{6.18}$$

As already noted before, if discount factors are less than or equal to one, interest rates (either continuously or discretely compounded) are nonnegative.

Example 6.5.6 Consider a bond market with only one future time instant, say after one year, and with payoff vector and cashflow matrix given by

$$\pi_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \qquad F_2 = \begin{bmatrix} 1 \\ 1 + r_L \\ -(1 + r_B) \end{bmatrix}$$

where all contracts can only be held in nonnegative quantities. The above specification means that cash can be kept safely, that the discretely compounded interest rate on a savings account is given by r_L, and that you can borrow money at the discretely compounded rate r_B. No a-priori assumptions are made on either r_L or r_B. According to Fact 6.5.3, in order to prevent arbitrage there must exist a number d that solves the following system of inequalities:

$$d > 0, \quad 1 \geq d, \quad 1 \geq (1 + r_L)d, \quad -1 \geq -(1 + r_B)d.$$

A general strategy for finding out whether such a system of inequalities is solvable is to write all inequalities as upper and lower bounds on the unknown d, and then to see whether the maximum of the upper bounds is equal than or at most equal to the minimum of the lower bounds, or equivalently whether each upper bound is less than or at most equal to each of the lower bounds.[7] A complication in this case is however that there are parameters in the system (namely r_L and r_B), and that the transformation to upper bounds and lower bounds for d depends on whether $1+r_B$ and $1+r_L$ are positive, negative, or zero. Therefore

[7] The same idea can be applied when there are several unknowns; the unknowns are then eliminated one by one. This technique is known as Fourier-Motzkin elimination.

in principle there are nine cases to consider. Of course the "normal" situation is the one in which both $1+r_B$ and $1+r_L$ are positive. In this case the inequalities above become

$$0 < d, \quad \frac{1}{1+r_B} \le d, \quad d \le 1, \quad d \le \frac{1}{1+r_L}\,.$$

A solution exists if and only if the four following inequalities are satisfied (compare all upper bounds to all lower bounds):

$$0 < 1, \quad 0 < \frac{1}{1+r_L}\,, \quad \frac{1}{1+r_B} \le 1, \quad \frac{1}{1+r_B} \le \frac{1}{1+r_L}\,.$$

The first inequality is always satisfied, and the second inequality is satisfied as well under the assumption that $1+r_L$ is positive. The remaining two inequalities can be translated into

$$r_B \ge 0, \qquad r_L \le r_B. \tag{6.19}$$

Among the other eight cases, the ones in which $1+r_B$ is zero or negative do not give rise to any solutions. In the cases in which $1+r_L$ is zero or negative, it turns out that only the equation $r_B \ge 0$ needs to be satisfied to avoid arbitrage. The set of parameter values (r_B, r_L) for which there is absence of arbitrage is therefore described by the conditions

$$1+r_B \ge 0 \text{ and } \big([1+r_L > 0 \text{ and } r_B \ge 0 \text{ and } r_L \le r_B] \text{ or } [1+r_L \le 0 \text{ and } r_B \ge 0]\big).$$

The condition on the left is redundant, and the remaining conditions can be merged so that in the end just the two conditions (6.19) remain. It can be verified directly that, if either of the conditions is not satisfied, it is possible to construct an arbitrage. The analysis shows that also the converse holds: when both conditions are satisfied, then no arbitrage is possible.

6.6 Exercises

1. Describe the situation of Example 5.1.3 in terms of a price vector π and a payoff matrix F. Give the arbitrage strategy in terms of a vector of holdings.

2. Repeat Exc. 1 for Examples 5.1.4–5.1.7, and for the example of Exc. 5.3.1.b. In each case, either demonstrate the presence of arbitrage by means of a vector of holdings representing an arbitrage strategy, or demonstrate the absence of arbitrage by means of Stiemke's lemma. Discuss the interpretation of positive solutions z of the matrix-vector equation $\pi = Fz$ in each of the examples.

3. Define a square matrix A by

$$A = \begin{bmatrix} a_1 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & a_2 & -1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & a_{n-1} & -1 \\ -1 & 0 & \cdots & \cdots & 0 & a_n \end{bmatrix}.$$

Show that

$$\det A = a_1 a_2 \cdots a_n - 1.$$

Can you relate this matrix fact to detection of arbitrage opportunities?

4. Suppose that the following is a payoff table for two government bonds.

	year 1	year 2
Bond 1	6	106
Bond 2	4	104

a. Let the price of the first bond be denoted by π_1, and the price of the second bond by π_2. In the (π_1, π_2)-plane, sketch the set of all pairs of prices (π_1, π_2) that are consistent with absence of arbitrage in the market defined by the two bonds.

b. Also indicate the set of all pairs of prices that are consistent with absence of arbitrage as well as with nonnegativity of interest rates.

5. Suppose that the following is a list of government bonds that are currently trading.

	Bond 1	Bond 2	Bond 3
Face value	100	100	100
Time to maturity	1 yr.	2 yr.	2 yr.
Coupon	4%	5%	2%
Current price	101	105	

Determine the arbitrage-free market value of Bond 3.

6. Consider a market in which there exist three default-free bonds as given in the following table.

	Bond 1	Bond 2	Bond 3
Face value	100	100	100
Time to maturity	3 yr.	2 yr.	3 yr.
Coupon	0%	4%	4%
Current price	x	100	100

Determine all values of x for which the market given by this table does not admit arbitrage opportunities.

7. **a**. Suppose that the current price of a three-year bond with 3% annual coupon rate is 100, and that the price of a three-year bond with 4% annual coupon rate is 102.90. Compute the bounds that are imposed by absence of arbitrage on the prices of the following products:

(i) a cashflow that generates 10 in the first year, 20 in the second year, 50 in the third year

(ii) a zero-coupon two-year bond with face value 100

(iii) a zero-coupon three-year bond with face value 100.

b. Can you explain the differences in the widths of the arbitrage-free price intervals for the three products?

8. Alice says: "If in the market there is a coupon-paying bond that trades at par, then the current value of a zero-coupon bond of the same maturity can at most be equal to its face value discounted at the par rate, or otherwise there is arbitrage." In more mathematical language, what she is saying is: suppose that at the price 1 one can buy a contract that pays r after 1, 2, $\dots$, $k-1$ years, and that pays $1+r$ after k years. Then the price of a contract that pays only 1 after k years can at most be equal to $(1+r)^{-k}$, or otherwise there is arbitrage. To investigate whether Alice is right, let us take $k=3$.

a. Consider a market in which there exist two default-free bonds as given in the following table.

	Bond 1	Bond 2
Face value	1	1
Time to maturity	3 yr.	3 yr.
Coupon	$100r$ %	0%
Current price	1	x

For any given positive value of r, determine for which values of x the market is free of arbitrage.

b. In cases in which the price x of the zero-coupon bond is so high that arbitrage is possible, construct the arbitrage strategy explicitly.

c. Is Alice's statement correct?

9. Find examples of all fourteen cases listed in Table 6.2.

6.7 Appendix: a proof of Farkas' lemma

Farkas' lemma states the following: for any given matrix $A \in \mathbb{R}^{m \times n}$ and vector $b \in \mathbb{R}^m$, either there exists a vector $x \in \mathbb{R}^m$ such that $x'A \geq 0$ and $x'b < 0$, or there exists a vector $z \in \mathbb{R}^n$ with $z \geq 0$ such that $b = Az$. When the matrix A is interpreted as a cashflow matrix and b as a corresponding price vector, then the statement can be rephrased equivalently as follows: there does not exist a vector x such that $x'A \geq 0$ and $x'b < 0$ (in financial terms, there does not exist a strict arbitrage) if and only if there does exist a vector $z \geq 0$ such that $b = Az$ (i.e., the prices of cashflows are obtained linearly from nonnegative prices of elementary securities). The "if" part is easy to show; if there exists $z \geq 0$ such that $b = Az$, then for any x such that $x'A \geq 0$ we have $x'b = x'Az \geq 0$. The "only if" part is more difficult. Let us state this part as a theorem. The proof of the theorem will be given after first a number of auxiliary results have been proved which will be stated as lemmas. The arguments used in the proofs are mathematical, but financial interpretations will be given throughout, since the line of reasoning becomes more natural in the light of these interpretations. For the sake of generality (since Farkas' lemma has many applications, not only in finance), statements and proofs will be phrased in the usual and more or less colorless terms of mathematics. However in financial interpretations the combination of a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$ will be referred to as a *market* with *cashflow matrix* A and *price vector* b, and the premise of the theorem below will be called *absence of strict arbitrage*.

Theorem 6.7.1 *Let a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$ be given, and suppose that the inequality $x'b \geq 0$ holds for any vector x such that $x'A \geq 0$. Then there exists a vector $z \geq 0$ such that $b = Az$.*

First let us show that, without loss of generality, we can assume that the rows of the matrix A are independent. In financial terms, the following lemma states that redundant assets must be priced conformably or else there is strict arbitrage.

Lemma 6.7.2 *Consider a matrix $A \in \mathbb{R}^{m\times n}$ and a vector $b \in \mathbb{R}^m$ with*

$$A = \begin{bmatrix} \tilde{A} \\ \hat{a} \end{bmatrix} \in \mathbb{R}^{(m+1)\times n}, \quad b = \begin{bmatrix} \tilde{b} \\ \hat{b} \end{bmatrix} \in \mathbb{R}^{m+1}$$

and suppose that the row vector $\hat{a}$ can be written as a linear combination of the rows of $\tilde{A}$. If $x'b \geq 0$ for all x such that $x'A \geq 0$, and if there exists $z \in \mathbb{R}^n$ such that $\tilde{b} = \tilde{A}z$, then the equality $\hat{b} = \hat{a}z$ holds as well.

Proof It is assumed that the row vector $\hat{a}$ is a linear combination of the rows of the matrix A, so there exists a vector $\tilde{x} \in \mathbb{R}^m$ such that $\hat{a} = \tilde{x}'\tilde{A}$. Then the vector $x' = [\tilde{x}' \;\; -1]'$ is such that $x'A = 0$. It follows that $x'b \geq 0$, but also $-x'b \geq 0$ since we have $x'A \geq 0$ as well as $-x'A \geq 0$; so in fact $x'b = 0$. This means that $\tilde{x}'\tilde{b} = \hat{b}$. Since $\hat{a}z = \tilde{x}'\tilde{A}z = \tilde{x}'\tilde{b}$, it follows that $\hat{b} = \hat{a}z$. □

In the proof of the theorem, we will be able on the basis of the above lemma to remove dependent rows from the matrix A. If we can find a vector z that satisfies the equation $b = Az$ for the reduced matrix A and the correspondingly reduced vector b, then the lemma guarantees that the same relation also holds for the original matrix and the original vector.

First consider now the case in which the rows of the matrix A span the space $\mathbb{R}^n$. In financial terms, this is the "complete market" case. Using Lemma 6.7.2, we can assume that the matrix C is invertible. The equation $b = Az$ then has a unique solution. If one or more of the prices of elementary securities is negative, it is easy to construct a strict arbitrage.

Lemma 6.7.3 *Consider an invertible matrix $A \in \mathbb{R}^{n\times n}$ and a vector $b \in \mathbb{R}^n$. If $x'b \geq 0$ for all x such that $x'A \geq 0$, then the vector z defined by $z = A^{-1}b$ satisfies $z \geq 0$.*

Proof Take $i \in \{1, \ldots, n\}$; let e_i denote the i-th unit vector. The vector x defined by $x' = e_i'A^{-1}$ is such that $x'A \geq 0$, so that by the assumption in the lemma $x'b \geq 0$. This implies $x'Az \geq 0$, i.e. $e_i'z \geq 0$. Consequently, $z_i \geq 0$ for all i, or in other words $z \geq 0$. □

Now consider the general case in which the rows of the matrix A may not span the n-dimensional space. Given the pair (A, b) and a row vector $\hat{a}$ of length n, define

$$p_*(\hat{a}) = \sup\{x'b \mid x'A \leq \hat{a}\}, \qquad p^*(\hat{a}) = \inf\{x'b \mid x'A \geq \hat{a}\}. \tag{6.20}$$

In financial terms, these are respectively the highest price that you can get by selling assets out of a payoff vector that you own, and the lowest price that you must pay in order to generate the same payoff vector. The following statement says that there is a strict arbitrage opportunity if there is a payoff vector that you can sell at a price that is higher than what you need to buy it.

Lemma 6.7.4 *Consider a matrix $A \in \mathbb{R}^{m\times n}$, a vector $b \in \mathbb{R}^m$, and a row vector $\hat{a}$ of length n. If $x'b \geq 0$ for all x such that $x'A \geq 0$, then $p_*(\hat{a}) \leq p^*(\hat{a})$.*

Proof Suppose that $p_*(\hat{a}) > p^*(\hat{a})$. Then there exist vectors x_1 and x_2 such that $x_1'A \leq \hat{a}$, $x_2'A \geq \hat{a}$, and $x_1'b > x_2'b$. But then $(x_2 - x_1)'b < 0$ while $(x_2 - x_1)'A \geq 0$, so that we have a contradiction. □

Theorem 6.7.1 will be proved by carrying out a stepwise algorithm. The steps are formulated in the lemma below. The financial meaning of the lemma is that, if a market is free of strict arbitrage and a new asset is introduced into the market, then it is always possible to find a price for the new asset so that the extended market is still free of strict arbitrage.

Lemma 6.7.5 *Let a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$ be given, and suppose that $x'b \geq 0$ for all $x \in \mathbb{R}^m$ such that $x'A \geq 0$. Let $\hat{a}$ be a row vector of the same length as the rows of A, and let $\hat{b}$ be chosen such that $p_*(\hat{a}) \leq \hat{b} \leq p^*(\hat{a})$ (which is possible by the previous lemma). Then the matrix $\tilde{A}$ and the vector $\tilde{b}$ defined by*

$$\tilde{A} = \begin{bmatrix} A \\ \hat{a} \end{bmatrix}, \qquad \tilde{b} = \begin{bmatrix} b \\ \hat{b} \end{bmatrix}$$

satisfy the property that $\tilde{x}'\tilde{b} \geq 0$ for all $\tilde{x} \in \mathbb{R}^{m+1}$ such that $\tilde{x}'\tilde{A} \geq 0$.

Proof Suppose to the contrary there exists a vector $\tilde{x} = [x' \;\; \hat{x}] \in \mathbb{R}^{m+1}$ such that

$$x'A + \hat{x}\hat{a} \geq 0 \qquad \text{and} \qquad x'b + \hat{x}\hat{b} < 0. \tag{6.21}$$

The scalar $\hat{x}$ must be nonzero, because otherwise we would have $x'A \geq 0$ and $x'b < 0$. Suppose for instance that $\hat{x} < 0$ (the argument is similar in case $\hat{x} > 0$). Then we may as well suppose that $\hat{x} = -1$. The inequalities (6.21) can then be rewritten as $x'A \geq \hat{a}$ and $x'b < \hat{b}$. Since $p^*(\hat{a})$ is the infimum of all numbers that can be obtained as $x'b$ for some x such that $x'A \geq \hat{a}$, it would then follow that $\hat{b} > p^*(\hat{a})$, which contradicts the assumption. □

Finally we arrive at a proof of the theorem. The idea in financial terms is to extend a given incomplete market by one asset at a time, making sure that prices are chosen such that introduction of strict arbitrage opportunities is avoided. After a number of steps the market has become complete, and it is then straightforward to derive that the prices of all assets must be obtained as linear combinations of nonnegative prices of elementary securities.

Proof (of Thm. 6.7.1). First assume that the rows of the matrix A are independent. If the rows moreover span the n-dimensional space, then the matrix A is invertible and existence of a nonnegative vector z such that $b = Az$ follows from Lemma 6.7.3. If they do not span the n-dimensional space, then, according to Lemmas 6.7.4 and 6.7.5, an independent row can be added together with a corresponding additional element of the vector b, in such a way that the hypothesis of the theorem also holds for the extended versions of the matrix

A and the vector b. If the rows of the extended matrix still do not span the n-dimensional space, another extension can be made, until after $n - \mathrm{rk}(A)$ steps the rows of the extended matrix do span the n-dimensional space. Let $\hat{A}$ denote the matrix consisting of rows that have been added, and let the elements added to the vector b be denoted by $\hat{b}$. Since we are now in the case of an invertible matrix, there is a nonnegative vector z such that

$$\begin{bmatrix} b \\ \hat{b} \end{bmatrix} = \begin{bmatrix} A \\ \hat{A} \end{bmatrix} z.$$

In particular, this implies $b = Az$.

When the rows of the matrix A are not independent, we can write (possibly after renumbering rows, which is immaterial)

$$A = \begin{bmatrix} \tilde{A} \\ \hat{A} \end{bmatrix}, \qquad b = \begin{bmatrix} \tilde{b} \\ \hat{b} \end{bmatrix}$$

where the rows of $\tilde{A}$ are independent, and the rows of $\hat{A}$ depend linearly on those of $\tilde{A}$. From the fact that $x'b \geq 0$ for all x such that $x'A \geq 0$, it follows in particular that $[\tilde{x} \;\; 0]'b \geq 0$ for all $\tilde{x}$ such that $[\tilde{x} \;\; 0]'A \geq 0$, which means that $\tilde{x}'\tilde{b} \geq 0$ for all $\tilde{x}$ such that $\tilde{x}'\tilde{A} \geq 0$. By the first part of the proof, there exists a vector z such that $\tilde{b} = \tilde{A}z$. Applying now Lemma 6.7.2 to the rows of $\hat{A}$, we can conclude that $b = Az$. $\square$

Chapter 7

Consistent pricing of risky assets

7.1 Objective, subjective, and implied probabilities

English-Dutch vocabulary for Section 7.1
implied probability geïmpliceerde kans

In the case of deterministic cashflows, we have usually considered *multiple* future times at which payments may take place. To analyze stochastic cashflows, we start with only *one* future time. This future date will be called time 1; current time is time 0. Instead of having multiple *dates*, we now assume that there are multiple *outcomes* that may arise at time 1. To keep things simple, we will assume that the number of possible different outcomes is finite, say n. For instance, if you toss a coin, there are two different outcomes ("heads" or "tails"). The uncertainty can be described by assigning *probabilities* $p_1, \ldots, p_n$ to the n outcomes that may arise at time 1.

The term "probability" will be used below in several different senses which should be carefully distinguished. First of all there is the notion of *objective probability*. This is the notion that is typically used in examples that appear in textbooks on probability theory: the probability of heads when a fair coin is tossed is 0.5, the probability of drawing a blue ball from an urn filled with 20 blue balls and 30 red balls is 0.4, and so on. In real life, one often works with *subjective probabilities*: the probability of failing in tomorrow's exam if I now go out and have a beer, the probability that the European Central Bank will increase its rates in the next board meeting, etcetera. These are called subjective probabilities, because they are based on personal judgment. Different individuals may assign different subjective probabilities to the same event.

To illustrate the notion of *implied probability*, consider the simple model of a defaultable loan that is shown in Table 7.1. When a firm defaults the creditors usually still can recover part of what the company owes them; for that reason the payoff in case of default is set at 52 rather than 0. Suppose that the discretely compounded interest rate for the maturity

price C	payoff in case of no default	payoff in case of default
99	104	52

Table 7.1: A simple model of a defaultable loan.

corresponding to time 1 is 4%. The price of the contract is denoted by C. The formula

$$C = \frac{1}{1.04}\big((1-p)\cdot 104 + p\cdot 52\big) \tag{7.1}$$

expresses the price of the contract as the expected discounted value of the payoff, if the quantity p represents the probability of default. Actually, the objective probability of default of a particular company at a particular time is not very easy to establish, given that it is not possible to repeat the same situation many times and to count in how many cases a default occurs. However the market price of the contract may be available because the contract may be traded, and then we can use the formula above to determine p when C is given. The quantity p that is obtained in this way is called an implied probability. In the example of Table 7.1 it is assumed that $C = 99$. From this it follows that $p = 0.02$, since $\frac{1}{1.04}(0.98 \cdot 104 + 0.02 \cdot 52) = 99$. Therefore we can say that the implied probability of default is 2%.

The implied probability is neither an objective nor a subjective probability. Neither is it correct to say that the implied probability is "the market's best estimate" of the actual probability. The purpose of the market is not to produce estimates that are optimized according to the criteria of statistics, but rather to reach equilibrium between demand and supply. Of course this equilibrium is influenced by the (subjective) probabilities that agents in the economy assign to the event of a default, but it is influenced by other factors as well such as in particular the degree of risk aversion of these agents. As discussed in Chapter 3, the implied probabilities are biased with respect to objective probabilities in the direction that is *un*favorable to the average investor; that is, bad events receive higher implied probability and good events receive lower implied probability. Therefore, implied probabilities are also called *risk-adjusted probabilities*.

In general terms, implied probabilities are defined (in the context of contract valuation) as follows. The implied probabilities of future events are the probabilities such that, when expectation is computed with respect to these probabilities of events, the prices computed as expected discounted values of future payoffs agree with the prices at which actual trades take place. Some authors have described implied probabilities as "the wrong numbers that you have to put into the wrong formula in order to get the right answers". In a sense that description is in fact correct, although it does not quite do justice to the important role that implied probabilities play in asset pricing, as will be discussed later on in this chapter.

In addition to the term "risk-adjusted probabilities" that was already mentioned above, also the term *risk-neutral probabilities* is sometimes used to describe implied probabilities. This is based on the idea that, in a risk-neutral world (that is, a world in which investors do not care about risk and only base their decisions on expected values corresponding to objective probabilities), there would be no difference between implied probabilities and objective probabilities.

7.2 Single-period market

English-Dutch vocabulary for Section 7.2	
exercise uitoefenen (van een optie)	**state** toestand

7.2.1 Definitions and terminology

Consider a single-period model in which n different situations may arise at time 1. These situations are called *states*. A financial product (asset) in this model is fully determined by giving its price at time 0 as well as its payoffs in all possible states at time 1. A model in which several assets are given in this way is called a *single-period market*. The assets whose payoffs and prices are given are called *basic assets*. A single-period market can be specified by a table in which the prices and payoffs of the basic assets are displayed; see Table 7.2 for an example with $n = 2$ (two states) and two basic assets. The data in the table can also be collected in a *price vector* and a *payoff matrix*; for instance in the case of Table 7.2 these are given by

$$\pi = \begin{bmatrix} 100 \\ 100 \end{bmatrix}, \quad G = \begin{bmatrix} 110 & 90 \\ 102 & 102 \end{bmatrix}. \tag{7.2}$$

The situation is analogous to the case of deterministic cashflows, except that the columns of the payoff matrix do not correspond to different *times*, but to different *states* that may arise at the same point in time. For instance, a row $[5 \;\; 105]$ in a cashflow matrix means that you will get *both* 5 and 105, be it at different times, but a row $[90 \;\; 110]$ in a payoff matrix means that you will get *either* 90 or 110, depending on the state that will materialize.

Given a single-period market, we can consider products whose payoffs are specified but whose prices at time 0 are not given (derivative products). We can ask what can be said about the prices of such products on the basis of the principle of absence of arbitrage. If there is a uniquely defined price, then we can also call it the market-consistent price of the derivative, where the market is defined by the prices of the basic assets. Market-consistent pricing of derivatives is the main topic of this chapter.

product	price	payoff in *up* state	payoff in *down* state
stock	100	110	90
bond	100	102	102

Table 7.2: A one-period model with two states at time 1.

The term "price" has up to now been used only for the value of an asset at time 0. Later on we will consider also multiperiod models and then the price of an asset is also relevant at later times. Even in a single-period model, the payoff at time 1 can also be considered as a price at time 1. Conversely, the price of an asset at some point in time in a multiperiod model is by definition the amount of cash for which the asset might be sold at that point, so that the price can be thought of as a potential payoff. This identification of "price" and "payoff" will be important in particular later when we discuss multiperiod models.

7.2.2 Assumptions

To formulate conditions for absence of arbitrage in a single-period market, first consider the following assumptions.

Assumption 7.2.1 Portfolios can be formed consisting of arbitrary linear combinations of products.

This is a liquidity assumption. In the context of Table 7.2, the assumption means that it is possible for instance to form a portfolio consisting of 3.14159 stocks and $-\sqrt{2}$ bonds, and that the price of that portfolio is $314.159 - 100\sqrt{2}$. The assumption that arbitrary numbers can be used, rather than only integers, does not really hold in practice but for most purposes there is little harm in assuming infinite divisibility of assets. To assume that *negative* holdings are possible is more serious. This part of the assumption might be dropped, as discussed in Section 6.5, but here we will hold on to it for reasons of simplicity.

Assumption 7.2.2 Agents agree that all states have positive probability.

This is partly a technical assumption and partly an efficiency assumption. We are not concerned about payoffs in impossible events, since such events have no effect on prices. Presumably nobody will pay anything for an insurance contract that protects against shark attack during a stay in the Himalaya. If agents differ in their assessments as to whether a state is possible or not, then arbitrage opportunities may arise. To avoid complications relating to this, we assume that there is agreement between the agents in the economy about which future states are possible. In other words, although the subjective probabilities assigned to the states may be different for different agents, there are no states which have positive probability according to some agents and zero probability according to other agents. In this way, the notion of "positive probability" that is used in the definition of arbitrage (Def. 5.1.1) is defined unambiguously.

Assumption 7.2.3 At least one of the products has positive price and nonnegative payoffs in all states.

This assumption is not really necessary, but it simplifies the formulation of the definition of arbitrage. Anyway, it is hard to think of a reasonable financial model that does not satisfy the assumption. A slightly stronger but still very reasonable assumption is the following.

Assumption 7.2.4 At least one of the products has positive price and positive payoffs in all states.

7.2.3 Arbitrage

Under the above assumptions, the general definition of arbitrage as given in Def. 5.1.1 can be made more specific for the single-period market as follows. In this definition, a "portfolio" is simply a combination of the basic assets that are available in the market (cf. Assumption 7.2.1).

Definition 7.2.5 The single-period market with n states at time 1 *allows arbitrage* if there exists a portfolio that has zero price, nonnegative payoffs in all n states, and a positive payoff in at least one state.

In other words: in an arbitrage opportunity you invest *nothing*, you have *zero* probability of a *negative* payoff, and you have *positive* probability of a *positive* payoff.

There may be a situation in which a product can be formed that has a *negative* price and nonnegative (possibly zero) payoffs in all future states. This represents an arbitrage opportunity according to Def. 5.1.1. Under Assumptions 7.2.1 and 7.2.3, it is always possible to convert such an arbitrage opportunity to one of the forms described in Def. 7.2.5. For instance, suppose we have the following situation.

product	price	payoff in state 1	payoff in state 2
R	-0.1	0	0
S	10	0	20

Product R represents an arbitrage opportunity in the sense of Def. 5.1.1, but not an arbitrage as described in Def. 7.2.5. However, due to the presence of product S which is of the form described in Assumption 7.2.3, we can form the portfolio consisting of one unit of asset R and 0.01 units of asset S. This portfolio has price 0, has nonnegative payoffs in all states, and has a positive payoff, namely 0.2, in state 2. This method works in general. Therefore, there exists an arbitrage opportunity in the sense of Def. 7.2.5 if and only if there exists an arbitrage opportunity in the sense of Def. 5.1.1. In summary, under Assumption 7.2.3, the two definitions come down to the same thing.

7.2.4 Conditions for absence of arbitrage

To write the condition for existence of an arbitrage opportunity in mathematical terms, we can use Def. 5.1.1 directly. Let the market be given by a price vector π and a payoff matrix G. Then there is an arbitrage opportunity in the market if and only if there exists a vector x such that

$$x'[-\pi \;\; G] \geq 0, \quad x'[-\pi \;\; G] \neq 0. \tag{7.3}$$

The condition is exactly the same as (6.1); the difference is just in the interpretation of the matrix G. The columns of G now correspond to different *states*, rather than to different *times*. To give a condition for absence of arbitrage, we can therefore copy Fact 6.2.2, except that the discount factors that appear in this statement should now get a different name since they are now factors that relate to different states. The "state factors" will be called *state prices*, denoted by π_s, because they are prices of state contracts. The condition for absence of arbitrage in the single-period market with finitely many future states can be given as follows.

Fact 7.2.6 *In a single-period market with finitely many future states, characterized by a price vector π and a payoff matrix G, there is absence of arbitrage if and only if there exists a vector π_s of positive state prices such that $\pi = G\pi_s$.*

Consider for instance the market defined in (7.2). In this case the matrix G is square and invertible so that there is only one vector π_s that satisfies the equation $\pi = G\pi_s$, namely

$$\pi_s = G^{-1}\pi = \begin{bmatrix} 0.5882 \\ 0.3922 \end{bmatrix}. \tag{7.4}$$

The entries of this vector are positive and so we can conclude that the market defined by (7.2) is free of arbitrage.

7.2.5 Complete markets

In analogy with Fact 6.3.1, we can state the following.

Fact 7.2.7 *The following conditions are equivalent for a payoff matrix G of size $m \times n$, corresponding to a single-period market with m assets and n future states:*

(i) *state prices are unique in the sense that for any given price vector $\pi \in \mathbb{R}^m$, there exists at most one vector $\pi_s \in \mathbb{R}^n$ such that $G\pi_s = \pi$;*

(ii) *all products can be replicated; in other words, for each row vector $\hat{g}$ of length n there exists a vector x of length m such that $\hat{g} = x'G$.*

A single-period market in which the above conditions hold is called, just as in the case of deterministic cashflows, a *complete market*. The elements of the (not necessarily uniquely determined) vector x that satisfies the equation $\hat{g} = x'G$ are called *replication weights*. A typical situation in which a complete market is defined is the one in which the number of basic assets m is equal to the number n of future states, and the matrix G is invertible. In this case the replication weights are determined uniquely.

The market defined by (7.2) is an example of a complete market. This means in particular that we can replicate all *derivative products* whose payoffs at time 1 are determined in terms of the payoffs of the basic assets which in this case are the stock S and the bond B. Take for instance a call option on one unit of the stock, with strike price 105. The call option gives the owner the right, but not the obligation, to buy one unit of the stock at time 1 for the price of 105 euro. This right will of course not be exercised at the down state, because it would be irrational to pay 105 euro for an item that can be bought for 90 euro. On the other hand, in the up state the option is worth 5 euro (the excess of the actual price, 110 euro, over the strike price, 105 euro). So the value of the option is nothing in the down state, and 5 euro in the up state. Replicating this payoff is a matter of solving the equation

$$x' \begin{bmatrix} 110 & 90 \\ 102 & 102 \end{bmatrix} = \begin{bmatrix} 5 & 0 \end{bmatrix}$$

for the unknown 2-vector x. The solution is (up to four decimals)

$$x = \begin{bmatrix} 0.25 \\ -0.2206 \end{bmatrix}. \tag{7.5}$$

And indeed, we have

$$0.25 \cdot 110 - 0.2206 \cdot 102 = 4.9988, \quad 0.25 \cdot 90 - 0.2206 \cdot 102 = -0.0012$$

where the deviation from the numbers 5 and 0 could be made to disappear by determining the second component of x more accurately (its exact value is $-\frac{15}{68}$).

It follows that, in a complete and arbitrage-free market, the prices of derivative products are determined uniquely by the condition of absence of arbitrage. In other words, if we impose that there should be no arbitrage in the market, then, given the prices of the assets that define the market (such as the stock and the bond in Table 7.2), the prices of all derivative products are uniquely determined.

This is an important point and therefore the reasoning behind it is given explicitly below. To set the stage, let the market be given by a payoff matrix G of size $m \times n$ and a price vector π of length m. Under the assumption that the market is complete and arbitrage-free, there is a uniquely determined vector of state prices π_s such that $\pi = G\pi_s$. Consider now a derivative product, given by its payoff vector $\hat{g}$ (a row vector of length n) and price $\hat{\pi}$. The fact that there is only one value of $\hat{\pi}$ that prevents arbitrage can now be argued in two ways.

First argument. The unique price that prevents arbitrage is

$$\hat{\pi} = \hat{g}\pi_s. \tag{7.6}$$

Indeed, consider the market that arises by adding the product with payoff vector $\hat{g}$ and price $\hat{\pi}$ to the existing market. We get a new price vector $\tilde{\pi}$ of length $m+1$ and a new payoff matrix $\tilde{G}$ of size $(m+1) \times n$, defined as follows:

$$\tilde{\pi} = \begin{bmatrix} \pi \\ \hat{\pi} \end{bmatrix}, \quad \tilde{G} = \begin{bmatrix} G \\ \hat{g} \end{bmatrix}.$$

According to Fact 7.2.5, absence of arbitrage holds if and only if the matrix-vector equation $\tilde{\pi} = \tilde{G}\tilde{\pi}_s$ has a positive solution. This equation is equivalent to the set of two equations $\pi = G\tilde{\pi}_s$ and $\hat{\pi} = \hat{g}\tilde{\pi}_s$. We know that the first of these two equations has the unique solution π_s, so the equation $\tilde{\pi} = \tilde{G}\tilde{\pi}_s$ has a solution (which is then automatically positive) if and only if (7.6) holds.

Second argument. Let x be a vector of replication weights[1] for the derivative, so that

$$\hat{g} = x'G. \tag{7.7}$$

Then the unique price that prevents arbitrage is

$$\hat{\pi} = x'\pi. \tag{7.8}$$

This is seen as follows. If $\hat{\pi} > x'\pi$, then an arbitrage can be created by buying the replicating portfolio and selling the derivative. If $\hat{\pi} < x'\pi$, then there is arbitrage by buying the derivative and selling the replicating portfolio. So only if $\hat{\pi} = x'\pi$ there is a possibility of absence of arbitrage. Since the derivative can be replicated by the basic assets, any portfolio that can be formed from the basic assets and the derivative can also be formed from the basic assets alone, and if $\hat{\pi} = x'\pi$ both portfolios have the same price. So in this case there can indeed be no arbitrage in the extended market since there was no arbitrage in the original market.

[1]The term "weights" refers here to absolute numbers rather than to relative numbers, so that the sum of the weights is not necessarily equal to 1. Also, the weights can be negative.

The two reasonings each lead to their own pricing formula. As it should be, the two formulas produce the same result: since $\hat{g} = x'G$ and $\pi = G\pi_s$, we have

$$x'\pi = x'G\pi_s = \hat{g}\pi_s.$$

Correspondingly, there are two ways to compute the price of a derivative:

(i) first compute the state prices from the equation $\pi = G\pi_s$, then compute the price of the derivative by $\hat{\pi} = \hat{g}\pi_s$

(ii) first compute the replication weights from the equation $\hat{g} = x'G$, then compute the derivative price by $\hat{\pi} = x'\pi$.

The first method is more convenient when the prices of several derivatives are to be computed, since the state prices do not depend on the particular derivative to be considered so they need not be recomputed for each new derivative, in contrast to the replication weights. An additional argument of a practical nature is that the computation via the state prices is usually less sensitive to roundoff errors than the one via the replication weights. This is not to say that one should never compute replication weights. In fact these weights are very important in risk management, as will be discussed below in subsection 7.3.4 and section 8.1.2.

Example 7.2.8 Let us compute in two ways the price of a call option with strike 105 in the market given by (7.2). The payoff vector of the call is given by $\hat{g} = [5 \;\; 0]$. The state prices were already computed in (7.4) and we find

$$\hat{\pi} = \begin{bmatrix} 5 & 0 \end{bmatrix} \begin{bmatrix} 0.5882 \\ 0.3922 \end{bmatrix} = 2.9410.$$

The replication weights for this option have also already been computed and are given in (7.5). From the replication weights the price of the derivative is found as follows:

$$\hat{\pi} = \begin{bmatrix} 0.25 & -0.2206 \end{bmatrix} \begin{bmatrix} 100 \\ 100 \end{bmatrix} = 2.94.$$

Consider now another derivative, for instance a put option on one unit of the stock, with strike price 100. This contract gives the holder the right, but not the obligation, to sell one unit of the stock at time 1 to the counterparty at the price of 100 euro. The value of this contract is 10 when the price of the stock is 90, and the value is 0 when the stock price is 110. Therefore the payoff vector of the put is $[0 \quad 10]$. To compute the price of this contract, the state prices can be used again and we get the answer 3.922. The price can also be computed via the replication weights, but then we first need to calculate these weights for the new derivative.

In the example given by (7.2), the state price corresponding to the up state is 0.5882 and the state price of the down state is 0.3922. These numbers are positive, as it should be the case in a complete and arbitrage-free market, but we cannot interpret them as implied probabilities because they do not add up to 1. In the next section, however, implied probabilities will be introduced for single-period markets. The key point is that, since payoffs are taking place in the future, we should also include a discount factor.

7.2.6 Strict arbitrage

A *strict* arbitrage opportunity in a single-period market is characterized according to Def. 5.1.2 as a portfolio that has a negative price and nonnegative payoffs. Under Assumption (7.2.4), this is equivalent to the following.

Definition 7.2.9 The single-period market with n states at time 1 *allows strict arbitrage* if there exists a portfolio that has zero price and positive payoffs in all states.

So a strict arbitrage opportunity requires zero investment and brings a positive outcome *for sure*. This is clearly stronger than a guarantee of a nonnegative outcome with a positive probability of a positive outcome, which corresponds to the notion of arbitrage that was used in most of this section. Again we can directly copy a result from the previous chapter, with just a change of interpretation, to get a characterization of absence of strict arbitrage in single-period markets.

Fact 7.2.10 *In a single-period market with finitely many future states, characterized by a price vector π and a payoff matrix G, there is absence of strict arbitrage if and only if there exists a vector π_s of nonnegative state prices such that $\pi = G\pi_s$.*

Because strict arbitrage is a stronger notion than arbitrage, *absence* of strict arbitrage is a *weaker* notion than absence of arbitrage. That is to say, it is easier for a market to meet the conditions for absence of strict arbitrage than to satisfy the conditions for absence of arbitrage. This is seen from the statement above: state prices do not need to be strictly positive as in the case of absence of arbitrage — they may also be zero.

In a complete market, the equation $\pi = G\pi_s$ for the state prices has at most one solution. If there is no solution, then there is arbitrage and even strict arbitrage according to Fact 7.2.10. If there is a solution and this solution has one or more negative components, then again there is strict arbitrage. If there are no negative components but there are one or more zero components, then there is no strict arbitrage opportunity but there is a non-strict arbitrage opportunity. In particular, such an arbitrage opportunity is presented by the asset that pays 1 in states with zero state price and 0 in all other states. Finally, if all components of the solution are positive, then the market is arbitrage-free.

The verification of these properties is somewhat more complicated in incomplete markets because of the nonuniqueness of state prices. For instance, even if a solution is found

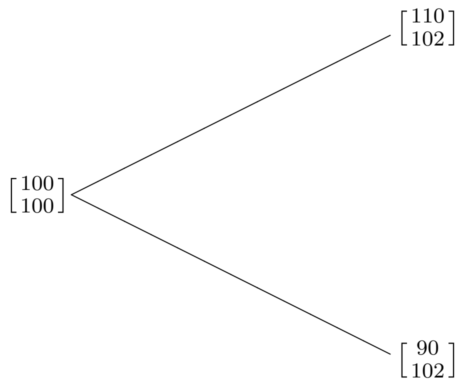

Figure 7.1: A one-step tree.

that is nonnegative with some zero components, it may be that there is another solution all of whose components are positive. In that case the market is free of arbitrage.

7.3 Pricing with implied probabilities

English-Dutch vocabulary for Section 7.3	
emanate from voortkomen uit	**node** knoop

7.3.1 Models with two states and two basic assets

A single-period market can be represented in tabular form, as in Table 7.2, or by the combination of a payoff matrix and a price vector, as in (7.2). Yet another presentation is the "tree form," which is given in Fig. 7.1 for the market defined by (7.2) and Table 7.2. This form becomes more interesting and useful in multiple-period markets. If as in the case shown in Fig. 7.1 there are two branches that emanate from each node except the end nodes from which there are no branches, then one speaks of a *binomial tree*. A single-period market with two future states is a one-step binomial tree. In a case such as in Fig. 7.1, where the values of one of the basic assets are the same in both end nodes, these values are often not written in the tree and are replaced by a separate specification of the interest rate.

Consider now a general binomial tree with two assets whose values are given both at time 0 and at time 1. The model may be represented in a tree form similar to Fig. 7.1, or in

product	current price	price in *up* state	price in *down* state
stock	S_0	S_u	S_d
bond	B_0	B_u	B_d
option	?	C_u	C_d

Table 7.3: General one-step tree in tabular form.

tabular form as in Table 7.3. For the sake of generality it is *not* assumed that $B_\mathrm{u} = B_\mathrm{d}$. We do assume though that the payoff matrix

$$G = \begin{bmatrix} S_\mathrm{u} & S_\mathrm{d} \\ B_\mathrm{u} & B_\mathrm{d} \end{bmatrix} \tag{7.9}$$

is invertible. If the two rows of G would be linearly dependent, then the payoffs of one of the assets at time 1 would be a fixed multiple of the payoffs of the other. In that case, one of the following situations must arise:

(i) the prices of the assets at time 0 relate in the same way as their payoffs; in that case one product is just a multiple of the other;

(ii) the prices do not relate in the same way as the payoffs; in that case the market is not arbitrage-free.

So if we want to consider an arbitrage-free market with two products that are not the same, we need to impose the invertibility assumption. Under this assumption, the single-period market defined by the two assets S and B is complete.

The invertibility of G is not enough to guarantee that the market is free of arbitrage. By Fact 7.2.6, the condition for this to hold is that the state prices π_u and π_d defined by

$$\begin{bmatrix} \pi_\mathrm{u} \\ \pi_\mathrm{d} \end{bmatrix} = G^{-1} \begin{bmatrix} S_0 \\ B_0 \end{bmatrix} \tag{7.10}$$

are positive.

Included in Table 7.3 is also a third asset, called “option”, whose values at both end nodes are given. We want to determine the unique value of the option at time 0 that is in accordance with the principle of absence of arbitrage. As discussed in the previous section, there are two ways of doing this. In terms of state prices, the option price is given by the formula $C_0 = C_\mathrm{u}\pi_\mathrm{u} + C_\mathrm{d}\pi_\mathrm{d}$. Alternatively we can use the formula $C_0 = x_S S_0 + x_B B_0$, where the replication weights are determined by

$$[x_S \quad x_B] = [C_\mathrm{u} \quad C_\mathrm{d}]G^{-1}. \tag{7.11}$$

The two formulas lead to the same result. Although we already have these two methods, we will discuss a third method which is based on implied probabilities. It is in some sense an improvement of the method of state prices; the advantages of implied probabilities become especially clear when later on we proceed to multistep trees.

In fact there are several ways to introduce implied probabilities. Below we first discuss what might be called the default method, and then a general method.

7.3.2 Pricing formula

Recall that the state prices (the components of the vector π_s) can be viewed as the prices of products that pay one euro in one of the states and nothing in all other states. It follows that the *sum* of the state prices is the price of the product that pays one euro in *all* states; in other words, this is the price of the product that pays one euro for sure. In bond market terminology, we have here a default-free zero-coupon bond maturing at time 1 with face value 1. It therefore makes sense to introduce a discretely compounded interest rate r for the period from time 0 to time 1, defined by

$$\frac{1}{1+r} = \pi_{\mathrm{u}} + \pi_{\mathrm{d}}. \tag{7.12}$$

Define

$$q_{\mathrm{u}} = (1+r)\pi_{\mathrm{u}}, \qquad q_{\mathrm{d}} = (1+r)\pi_{\mathrm{d}}. \tag{7.13}$$

The two numbers q_{u} and q_{d} are positive and their sum is 1, so they can be viewed as probabilities. Moreover, by definition of the state prices π_{u} and π_{d}, the relations $S_0 = S_{\mathrm{u}}\pi_{\mathrm{u}} + S_{\mathrm{d}}\pi_{\mathrm{d}}$ and $B_0 = B_{\mathrm{u}}\pi_{\mathrm{u}} + B_{\mathrm{d}}\pi_{\mathrm{d}}$ hold (cf. (7.10) and (7.9)). With use of the definitions (7.13), these relations can now be written as

$$S_0 = \frac{1}{1+r}(q_{\mathrm{u}}S_{\mathrm{u}} + q_{\mathrm{d}}S_{\mathrm{d}}), \qquad B_0 = \frac{1}{1+r}(q_{\mathrm{u}}B_{\mathrm{u}} + q_{\mathrm{d}}B_{\mathrm{d}}). \tag{7.14}$$

In other words, the current prices S_0 and B_0 can be interpreted as the expected discounted value of the future payoffs if r is taken as the one-period interest rate, and *if expectation is taken with respect to the probabilities* q_{u} *and* q_{d}. Therefore the numbers q_{u} and q_{d} can be interpreted as the *implied probabilities* of the up state and the down state respectively.

Some notation can be introduced to support the technique of computing prices as discounted expected values of future payoffs with respect to implied probabilities. Instead of (7.14), write

$$S_0 = \frac{1}{1+r}E^Q S_1, \qquad B_0 = \frac{1}{1+r}E^Q B_1 \tag{7.15}$$

where S_1 is the stochastic variable that takes the value S_{u} with probability q_{u} and the value S_{d} with probability q_{d}, B_1 is defined likewise, and the expectation symbol E is written with a mark Q to emphasize that the expectation is taken with respect to the implied probabilities q_{u} and q_{d} rather than to other probabilities, such as objective or subjective probabilities. Of course if $B_{\mathrm{u}} = B_{\mathrm{d}}$ then B_1 is a degenerate random variable, which means that it is not really random. The notation (7.15) is convenient because it is compact and can be used also in more general cases, for instance in multistep trees as will be seen later.

We already know that derivative products must be priced by the formula $C_0 = C_{\mathrm{u}}\pi_{\mathrm{u}} + C_{\mathrm{d}}\pi_{\mathrm{d}}$, or otherwise there is arbitrage. Introducing the stochastic variable C_1 that takes the value C_{u} with probability q_{u} and the value C_{d} with probability q_{d}, we can write (compare to (3.12)):

$$C_0 = \frac{1}{1+r} E^Q C_1. \tag{7.16}$$

In other words, *all products in the market are priced by the same formula*. More specifically, we can also say: *when expectation is taken with respect to implied probabilities, all assets have the same expected return* (namely r). Note that this motto is used in two ways: first to *define* the implied probabilities q_{u} and q_{d} and the implied interest rate r, and then, once these quantities have been determined, to compute prices of *derivatives*.

In summary, the pricing procedure can be described as follows. Consider a one-step binomial tree model with two assets, and suppose that this market is complete and arbitrage-free. First we need to find the implied interest rate and the implied probabilities of the two states. The three unknowns r, q_{u}, and q_{d} can be solved from the three equations

$$q_{\mathrm{u}}S_{\mathrm{u}} + q_{\mathrm{d}}S_{\mathrm{d}} = (1+r)S_0 \tag{7.17a}$$

$$q_{\mathrm{u}}B_{\mathrm{u}} + q_{\mathrm{d}}B_{\mathrm{d}} = (1+r)B_0 \tag{7.17b}$$

$$q_{\mathrm{u}} + q_{\mathrm{d}} = 1. \tag{7.17c}$$

Subsequently, whenever a contract is given by its payoffs in the two states at time 1, its arbitrage-free price at time 0 can be computed by the formula (7.16). This formula can be written more explicitly as

$$C_0 = \frac{1}{1+r}(q_{\mathrm{u}}C_{\mathrm{u}} + q_{\mathrm{d}}C_{\mathrm{d}}). \tag{7.18}$$

Example 7.3.1 Consider once more the market defined by (7.2). The three equations (7.17) become

$$q_{\mathrm{u}} \cdot 110 + q_{\mathrm{d}} \cdot 90 = (1+r) \cdot 100 \tag{7.19a}$$

$$q_{\mathrm{u}} \cdot 102 + q_{\mathrm{d}} \cdot 102 = (1+r) \cdot 100 \tag{7.19b}$$

$$q_{\mathrm{u}} + q_{\mathrm{d}} = 1. \tag{7.19c}$$

Because in this case the relation $B_{\mathrm{u}} = B_{\mathrm{d}}$ holds, the second equation simplifies to $102 = (1+r)100$, so that $r = 0.02$. Now rewrite the first equation as $110q_{\mathrm{u}} + 90(1-q_{\mathrm{u}}) = 102$; we find $q_{\mathrm{u}} = 0.6$ and hence $q_{\mathrm{d}} = 0.4$. The implied interest rate and the implied probabilities have been determined. Now we can price any product in the market, for instance the call option with payoff vector $[5 \;\; 0]$. According to formula (7.16), the price of the call option is given by

$$C_0 = \frac{1}{1.02}(0.6 \cdot 5 + 0.4 \cdot 0) = \frac{3}{1.02} = 2.9412$$

which is of course, up to rounding errors, the same as the result we found before by different methods.

7.3.3 General numéraire

The idea of the general method is that prices of assets can be expressed in units of another asset, rather than in euros. In fact, the price of an asset expressed in some unit of currency has no meaning by itself; only relative prices are important. So no information is lost if we express the value of assets in terms of units of some particular asset. If a particular asset is used as a unit of account in this way, it is called a *numéraire*. Any asset that has positive value at the current time as well as in all possible future states can be taken as a numéraire.

In the numéraire-based method, the price of a derivative product is obtained in terms of the numéraire. Conversion to euro is possible if we have the price of the numéraire in euros at time 0. This is the case if we define the numéraire as one of the basic assets, or as a combination of the basic assets.

Let the chosen numéraire be denoted by the letter M; its price at time 0 is M_0, and the payoffs in the up and down states are M_u and M_d respectively. It is assumed that the numéraire is itself a product that can be traded in the market without causing arbitrage, so that the following relation must hold:

$$M_0 = M_\mathrm{u}\pi_\mathrm{u} + M_\mathrm{d}\pi_\mathrm{d}. \tag{7.20}$$

Moreover, all three numbers M_0, M_u, and M_d are supposed to be positive. Now define

$$q_\mathrm{u} = \frac{M_\mathrm{u}}{M_0}\pi_\mathrm{u}, \qquad q_\mathrm{d} = \frac{M_\mathrm{d}}{M_0}\pi_\mathrm{d}. \tag{7.21}$$

These two numbers are positive and their sum is 1, because

$$q_\mathrm{u} + q_\mathrm{d} = \frac{M_\mathrm{u}\pi_\mathrm{u} + M_\mathrm{d}\pi_\mathrm{d}}{M_0} = 1$$

by (7.20). So we can view q_u and q_d as probabilities. The formula $C_0 = C_\mathrm{u}\pi_\mathrm{u} + C_\mathrm{d}\pi_\mathrm{d}$ which holds for all financial products can be written in terms of q_u and q_d as

$$\frac{C_0}{M_0} = q_\mathrm{u}\frac{C_\mathrm{u}}{M_\mathrm{u}} + q_\mathrm{d}\frac{C_\mathrm{d}}{M_\mathrm{d}}\,. \tag{7.22}$$

If we introduce M_1 as the stochastic variable that takes the value M_u with probability q_u and the value M_d with probability q_d, then we can also write more simply

$$\frac{C_0}{M_0} = E^{Q_M}\frac{C_1}{M_1}\,. \tag{7.23}$$

where the mark Q_M is attached to the expectation symbol as a reminder that expectation is taken with respect to implied probabilities which themselves depend on the choice of the numéraire. There is no explicit discounting factor in this formula, because the price of the derivative product is taken relative to the numéraire both at time 0 and at time 1. If we take a zero-coupon default-free bond maturing at time 1 as the numéraire, then $M_\mathrm{u} = M_\mathrm{d}$ so that we must have $M_\mathrm{u} = M_\mathrm{d} = (1+r)M_0$ where r is the implied interest rate, and the formula (7.23) reduces to (7.16) which does show the discounting explicitly. The numbers q_u and q_d defined by (7.21) are *not* the same as the ones defined by (7.13), except when $M_\mathrm{u} = M_\mathrm{d}$.

The implied probabilities corresponding to the numéraire M can be computed from the state prices via the definition (7.21). Since the pricing formula (7.23) applies to all assets and to the basic assets in particular, the implied probabilities can also be obtained more directly from the given asset prices and payoffs by solving the equations

$$\frac{S_0}{M_0} = q_\mathrm{u}\frac{S_\mathrm{u}}{M_\mathrm{u}} + q_\mathrm{d}\frac{S_\mathrm{d}}{M_\mathrm{d}} \tag{7.24a}$$

$$\frac{B_0}{M_0} = q_\mathrm{u}\frac{B_\mathrm{u}}{M_\mathrm{u}} + q_\mathrm{d}\frac{B_\mathrm{d}}{M_\mathrm{d}} \tag{7.24b}$$

$$q_\mathrm{u} + q_\mathrm{d} = 1 \tag{7.24c}$$

for the unknowns q_u and q_d. Although these are three equations in two unknowns, we know from the above that there is a solution, so the equations must be dependent. What kind of dependency exists between the equations depends on the choice of the numéraire.

Example 7.3.2 Again consider the market defined by (7.2). Let us take the stock S as a numéraire. Note that in this case the equations (7.24a) and (7.24c) are the same, so we solve for q_u and q_d from (7.24b) and (7.24c). The equations are

$$\frac{102}{110}q_\mathrm{u} + \frac{102}{90}q_\mathrm{d} = 1$$

$$q_\mathrm{u} + q_\mathrm{d} = 1$$

and the solution is $q_\mathrm{u} = \frac{11}{17} = 0.6471$, $q_\mathrm{d} = \frac{6}{17} = 0.3529$. The price of the call option with strike 105 is found from

$$\frac{C_0}{100} = 0.6471\tfrac{5}{110} + 0.3529\tfrac{0}{90}$$

which implies that $C_0 = 2.9414$. Up to rounding error, this is the answer we found before.

Since the pricing formula that we use gives the current *relative* price of any asset (relative to the numéraire) as the expected value of the *relative* payoff at time 1, the implied probabilities must depend on the numéraire. Otherwise the outcome of the formula would depend on the choice of the numéraire, which would be in disagreement with the fact that in a complete market the price of any derivative product is determined uniquely. The dependence of the implied probabilities on the numéraire may seem strange at first, but it just underlines the fact that these probabilities do not have an objective status. In the calculation, the implied probabilities are used together with the numéraire in such a way that the final result does not depend on the choice of the numéraire.

7.3.4 Replication recipe

The replication weights in a one-step binomial tree with two assets are given by the formula (7.11). If the option price C_0 at time 0 has already been computed from implied probabilities, then there are in fact even three equations for which the replication weights can be determined:

$$x_S S_\mathrm{u} + x_B B_\mathrm{u} = C_\mathrm{u}, \qquad x_S S_\mathrm{d} + x_B B_\mathrm{d} = C_\mathrm{d}, \qquad x_S S_0 + x_B B_0 = C_0.$$

By construction, the number C_0 is such that these equations are redundant, so that a solution still exists even though we have three equations for only two unknowns. From the first and the third equation,[2] we can write down the following explicit formulas:

$$x_S = \frac{C_\mathrm{u}/B_\mathrm{u} - C_\mathrm{d}/B_\mathrm{d}}{S_\mathrm{u}/B_\mathrm{u} - S_\mathrm{d}/B_\mathrm{d}} \tag{7.25a}$$

$$x_B = \frac{C_0 - x_S S_0}{B_0}\,. \tag{7.25b}$$

In the case in which $B_\mathrm{u} = B_\mathrm{d} = (1+r)B_0$, the formula for the stock holding x_S simplifies to

$$x_S = \frac{C_\mathrm{u} - C_\mathrm{d}}{S_\mathrm{u} - S_\mathrm{d}}\,. \tag{7.26}$$

Example 7.3.3 Consider again the market of Table 7.2 and the payoff vector $[0 \quad 10]$ that corresponds to a put option with strike 100. The price of this option was already computed before on the basis of the state prices that were obtained in (7.4); the result is $C_0 = 3.922$. According to formula (7.26), the number of units of the stock to be held in a replicating portfolio is given by

$$x_S = \frac{0 - 10}{110 - 90} = -0.5.$$

The number of units of the bond in the replicating portfolio can be computed from (7.25b):

$$x_B = \frac{3.922 + 0.5 \cdot 100}{100} = 0.53922.$$

The replication weights can be used not only for pricing but also for *risk management*. An institution that has written the call option (which means that at time 1 it has to sell the underlying to the holder of the option for the strike price) can *hedge* this liability by using the option premium to buy a hedging portfolio. The value of this portfolio at time 1 is exactly what is needed to make the payment that is required by the contract, both when the u event takes place and when the d event takes place.

The fact that replication is possible shows that the price is fixed by the market data. More precisely, it is enough to have the price scenarios of the assets, since these determine

[2]Computing the replication weights in this way ensures that the weights define a self-financing strategy, even when some mistake has been made in the state price method which has caused the option values to be computed incorrectly.

the implied probabilities from which the price of the option can be computed — equivalently, they determine the composition of the hedge portfolio by which the option can be replicated. In particular *no* information is needed concerning the objective probabilities of the future states u and d. As a consequence, the real-world expected return of the risky asset S is not relevant to the option pricing problem.

7.3.5 Multinomial one-step trees

So far we have considered one-step trees with two branches and two basic assets. We can also consider markets associated to one-step trees with three or more branches. To get a complete market in such cases, the number of independent basic assets should be equal to the number of branches. We can then compute the implied interest rate and implied probabilities for all states, and these can be used to find the arbitrage-free price of any derivative product. The procedure is illustrated in the following example.

Example 7.3.4 Consider a model in which there are three possible future states, labeled "up," "down," and "crisis." The basic assets are a stock, a bond, and a put option on the stock with strike 105. Because there are three future states, the arbitrage-free price of the put is not uniquely determined by the prices of the stock and the bond; therefore we take it as a basic asset. Suppose that the prices and payoffs of the three products are as in Table 7.4. The implied interest rate and the implied probabilities of the three states (according to the default method, which comes down to taking the bond as a numéraire) are determined from the equations

$$\begin{aligned} q_{\mathrm{u}} \cdot 126 + q_{\mathrm{d}} \cdot 84 + q_{\mathrm{c}} \cdot 42 &= (1+r) \cdot 96 \\ q_{\mathrm{u}} \cdot 105 + q_{\mathrm{d}} \cdot 105 + q_{\mathrm{c}} \cdot 105 &= (1+r) \cdot 100 \\ q_{\mathrm{u}} \cdot 0 + q_{\mathrm{d}} \cdot 21 + q_{\mathrm{c}} \cdot 63 &= (1+r) \cdot 14 \\ q_{\mathrm{u}} + q_{\mathrm{d}} + q_{\mathrm{c}} &= 1 \end{aligned}$$

where q_{c} denotes the implied probability of the crisis state. From the second and the fourth equation it readily follows that $r = 0.05$. Some further calculation shows that $q_{\mathrm{u}} = 0.5$, $q_{\mathrm{d}} = 0.4$, $q_{\mathrm{c}} = 0.1$. We can now for instance compute the arbitrage-free price of a call option with strike 100. The payoff vector of this option is $[26 \quad 0 \quad 0]$ and so its price is $C_0 = \frac{1}{1.05} \cdot 26 \cdot 0.5 = 12.3810$.

7.4 Exercises

1. Consider a single-period market with two assets and two states at time 1. The two future states are denoted by u and d; the assets are denoted by S and B. The current prices of the two assets are S_0 and B_0. The payoff of asset S in state u is $(1 + r_{\mathrm{u}})S_0$, while in state d it is $(1 + r_{\mathrm{d}})S_0$; here, r_{u} and r_{d} are numbers such that $r_{\mathrm{u}} > r_{\mathrm{d}}$. The payoff of asset B is equal

	current price	price in up state	price in down state	price in crisis state
stock	96	126	84	42
bond	100	105	105	105
put	14	0	21	63

Table 7.4: A market with three assets and three future states.

to $(1+r)B_0$ in both states. Find the conditions on the three numbers $r_{\mathtt{u}}$, $r_{\mathtt{d}}$, and r under which this market is free of arbitrage. Under which conditions is the market free of strict arbitrage?

Chapter 8

Tree models and option pricing

8.1 Two-step trees

English-Dutch vocabulary for Section 8.1
recombining recombinerend

8.1.1 Pricing in a two-step tree

In a complete single-period market, any given payoff vector is a linear combination of the payoff vectors of the basic assets. The arbitrage-free price of product corresponding to the given payoff vector is just the corresponding linear combination of the prices of the basic assets. To introduce implied probabilities for the purpose of pricing in such markets could be considered a case of overkill. However, it turns out that implied probabilities are very convenient in two-period and more generally multiperiod markets.

In a two-step binomial tree, there are two states u and d at time 1, and four states uu, ud, du, and dd at time 2. The states uu and ud can be reached from state u, and du and dd can be reached from state d. To form a two-step *market*, prices of basic assets should be given at all nodes of the tree, including the initial state at time 0. The tree is said to be *recombining* if there is no difference between asset prices at node du and at node ud, so that these two nodes together essentially form one state.

An example of a two-step tree with stock prices is given in Fig. 8.2. This tree is non-recombining. In the context of two-step trees it is often assumed that there is a fixed interest rate which is the same for the period from time 0 to time 1 and for the period from time 1 to time 2. Specifically, if the interest rate per period is 3%, then a second asset called "bond" is added in the tree of Fig. 8.2 which has value 100 at time 0, 103 in both nodes at time 1, and 106.90 in all four nodes at time 2. In other words, we have $B_0 = 100$, $B_{\mathrm{u}} = B_{\mathrm{d}} = 103$, $B_{\mathrm{uu}} = B_{\mathrm{ud}} = B_{\mathrm{du}} = B_{\mathrm{dd}} = 106.90$.

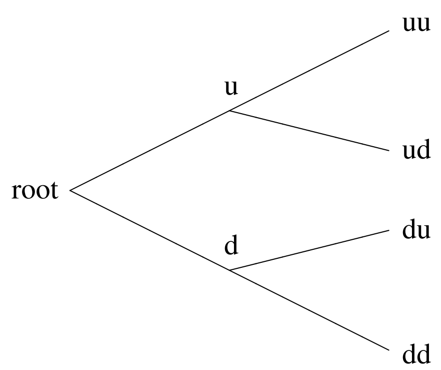

Figure 8.1: A two-step tree.

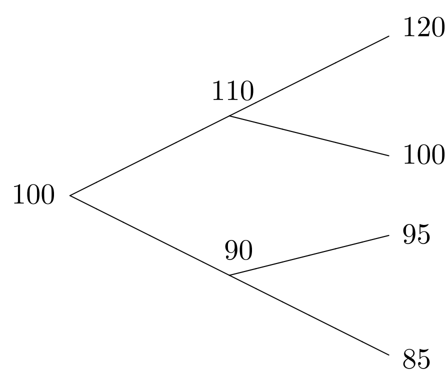

Figure 8.2: A two-period market.

The two-step tree is a *dynamic* model (it has multiple time steps), and therefore in the context of this model we can consider *dynamic strategies*. Whereas before a "strategy" consisted just of a decision concerning the portfolio composition, in the context of a dynamic model it is possible to take portfolio decisions at later times as well depending on information that has become available in the meantime. Specifically, in the context of a two-step tree, the portfolio composition may be revised at time 1 depending on whether the up event or the down event has materialized. We shall work under the following assumption.

Assumption 8.1.1 At each node, the portfolio composition can be changed arbitrarily subject to only one constraint, namely that the total value of assets in the portfolio before the change of composition must be the same as the total value of assets in the portfolio after the change of composition.

The assumption implies that a *budget constraint* is imposed. Stated briefly (and not very precisely), if you want to buy something, you have to sell something else. For instance, if you have two units of asset 1 with current value 100 and three units of asset 2 with current value 50, then you can sell two units of asset 2 and buy one unit of asset 1, so that after that you have three units of asset 1 and one unit of asset 2; the total value of assets is 350 both before and after the transaction. This also means that it is assumed that transactions take place instantaneously, so that prices have no time to change, and that transactions costs are ignored. A portfolio strategy that meets the budget constraint stated in the above assumption is said to be a *self-financing strategy*. Indeed, such a strategy finances itself in the sense that, after the portfolio has been formed at the initial time, no influx of funds is required under any circumstances, nor is any outflow of funds permitted.

In the context of the market defined in this way, consider now a derivative product expiring at time 2. Take for instance a put option on S with strike $K = 105$. The value of the put at time 2 is given by $\max(K - S_2, 0)$, or in tabular form:

C_{uu}	C_{ud}	C_{du}	C_{dd}
0	5	10	20

Consider the part of the tree that starts from node u. This part is a one-step tree and it represents a complete market with S_{u} and B_{u} as the initial values of the two basic assets. We can compute, by any of the methods discussed in the previous section, a number C_{u} that represents the value of a replicating portfolio at node u for the outcomes C_{uu} and C_{ud}. Likewise, at node d there is a number C_{d} that represents the value of a replicating portfolio for the outcomes C_{du} and C_{dd}. When we look at the initial node and look no further than time 1, we again have a one-step tree, and we can find a number C_0 that represents the value of a portfolio that is constructed in such a way that it has value C_{u} in node u and C_{d} in node d. This leads to a self-financing strategy that replicates the payoff table at time 2.

Details of the replication method are given in Section 8.1.2 below. To compute the arbitrage-free *price* of the option at time 0, it is not necessary to compute replication weights because we can use the method of state prices instead. The procedure is as follows.

Start by calculating the implied probabilities for the branches at all nodes. The implied probabilities of the u and d events at the "up" and the "down" node are indicated by q_{uu} and q_{ud} and by q_{du} and q_{dd} respectively, whereas at the root of the node the implied probabilities are denoted by q_{u} and q_{d} as before. The implied probabilities at the "up" node are found from

$$110 = \frac{1}{1.03}\big(120\, q_{\text{uu}} + 100\,(1 - q_{\text{uu}})\big)$$

and those at the "down" node from

$$90 = \frac{1}{1.03}\big(95\, q_{\text{du}} + 85\,(1 - q_{\text{du}})\big).$$

Calculation shows that

$$q_{\text{uu}} = 0.665, \quad q_{\text{ud}} = 0.335, \quad q_{\text{du}} = 0.77, \quad q_{\text{dd}} = 0.23.$$

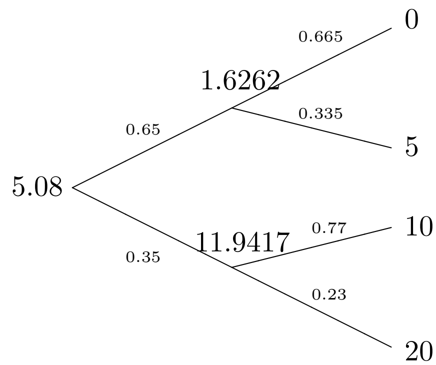

Figure 8.3: Option values in a two-step tree.

We can now compute C_{u} and C_{d}:

$$C_{\mathrm{u}} = \frac{1}{1.03}\big(0.665 \cdot 0 + 0.335 \cdot 5\big) = 1.6262$$

$$C_{\mathrm{d}} = \frac{1}{1.03}\big(0.77 \cdot 10 + 0.23 \cdot 20\big) = 11.9417.$$

To compute C_0, we now need the implied probabilities q_{u} and q_{d}. These are computed from

$$100 = \frac{1}{1.03}\big(110\, q_{\mathrm{u}} + 90\,(1 - q_{\mathrm{u}})\big).$$

We obtain $q_{\mathrm{u}} = 0.65$, $q_{\mathrm{d}} = 0.35$. The option price at time 0 is therefore given by

$$C_0 = \frac{1}{1.03}\big(0.65 \cdot 1.6262 + 0.35 \cdot 11.9417\big) = 5.08.$$

The calculation is summarized in Fig. 8.3.

Now, let's go through this procedure in general, following the method of Subsection 7.3.3.[1] Assume that a numéraire asset has been chosen. The corresponding implied probabilities are obtained from the equations

$$\frac{S_{\mathrm{u}}}{M_{\mathrm{u}}} = q_{\mathrm{uu}}\frac{S_{\mathrm{uu}}}{M_{\mathrm{uu}}} + q_{\mathrm{ud}}\frac{S_{\mathrm{ud}}}{M_{\mathrm{ud}}}, \quad \frac{B_{\mathrm{u}}}{M_{\mathrm{u}}} = q_{\mathrm{uu}}\frac{B_{\mathrm{uu}}}{M_{\mathrm{uu}}} + q_{\mathrm{ud}}\frac{B_{\mathrm{ud}}}{M_{\mathrm{ud}}}, \quad q_{\mathrm{uu}} + q_{\mathrm{ud}} = 1 \tag{8.1a}$$

$$\frac{S_{\mathrm{d}}}{M_{\mathrm{d}}} = q_{\mathrm{du}}\frac{S_{\mathrm{du}}}{M_{\mathrm{du}}} + q_{\mathrm{dd}}\frac{S_{\mathrm{dd}}}{M_{\mathrm{dd}}}, \quad \frac{B_{\mathrm{d}}}{M_{\mathrm{d}}} = q_{\mathrm{du}}\frac{B_{\mathrm{du}}}{M_{\mathrm{du}}} + q_{\mathrm{dd}}\frac{B_{\mathrm{dd}}}{M_{\mathrm{dd}}}, \quad q_{\mathrm{du}} + q_{\mathrm{dd}} = 1 \tag{8.1b}$$

$$\frac{S_0}{M_0} = q_{\mathrm{u}}\frac{S_{\mathrm{u}}}{M_{\mathrm{u}}} + q_{\mathrm{d}}\frac{S_{\mathrm{d}}}{M_{\mathrm{d}}}, \quad \frac{B_0}{M_0} = q_{\mathrm{u}}\frac{B_{\mathrm{u}}}{M_{\mathrm{u}}} + q_{\mathrm{d}}\frac{B_{\mathrm{d}}}{M_{\mathrm{d}}}, \quad q_{\mathrm{u}} + q_{\mathrm{d}} = 1. \tag{8.1c}$$

[1]The formulation in terms of a general numéraire makes it possible for instance to handle cases in which the interest rate in the second period depends on whether the up node or the down node is realized (stochastic interest rates).

These three sets of equations are independent of each other, so they can be solved in any order. Now determine C_{u} and C_{d} by

$$\frac{C_{\text{u}}}{M_{\text{u}}} = q_{\text{uu}}\frac{C_{\text{uu}}}{M_{\text{uu}}} + q_{\text{ud}}\frac{C_{\text{ud}}}{M_{\text{ud}}} \tag{8.2a}$$

$$\frac{C_{\text{d}}}{M_{\text{d}}} = q_{\text{du}}\frac{C_{\text{du}}}{M_{\text{du}}} + q_{\text{dd}}\frac{C_{\text{dd}}}{M_{\text{dd}}} \tag{8.2b}$$

and finally, determine C_0 from

$$\frac{C_0}{M_0} = q_{\text{u}}\frac{C_{\text{u}}}{M_{\text{u}}} + q_{\text{d}}\frac{C_{\text{d}}}{M_{\text{d}}}\,. \tag{8.2c}$$

This equation uses the values of C_{u} and C_{d}, so we need to compute these values first before we can compute C_0. The computation of the option price is a *backward* procedure.

Taking the equations (8.2) together, we can write the following formula which gives the option value directly in terms of the option payoffs and the implied probabilities:

$$\frac{C_0}{M_0} = q_{\text{u}}\left(q_{\text{uu}}\frac{C_{\text{uu}}}{M_{\text{uu}}} + q_{\text{ud}}\frac{C_{\text{ud}}}{M_{\text{ud}}}\right) + q_{\text{d}}\left(q_{\text{du}}\frac{C_{\text{du}}}{M_{\text{du}}} + q_{\text{dd}}\frac{C_{\text{dd}}}{M_{\text{dd}}}\right). \tag{8.3}$$

Note that q_{uu} and q_{ud} refer to the implied probabilities of reaching state uu and ud respectively at time 2, *given* that at time 1 state u has been reached. From the point of view of time 0, the probability of reaching state uu at time 2 is $q_{\text{u}}q_{\text{uu}}$ since we first need to reach state u at time 1 and from there we must reach uu at time 2. Write

$$q^0_{\text{uu}} = q_{\text{u}}q_{\text{uu}}, \quad q^0_{\text{ud}} = q_{\text{u}}q_{\text{ud}}, \quad q^0_{\text{du}} = q_{\text{d}}q_{\text{du}}, \quad q^0_{\text{dd}} = q_{\text{d}}q_{\text{dd}}.$$

These are the probabilities of the four states at time 2 as seen from time 0, which might be called the "unconditional" probabilities, i.e. the probabilities given only information at time 0. In terms of these probabilities, the pricing formula (8.3) can be rewritten as follows:

$$\frac{C_0}{M_0} = q^0_{\text{uu}}\frac{C_{\text{uu}}}{M_{\text{uu}}} + q^0_{\text{ud}}\frac{C_{\text{ud}}}{M_{\text{ud}}} + q^0_{\text{du}}\frac{C_{\text{du}}}{M_{\text{du}}} + q^0_{\text{dd}}\frac{C_{\text{dd}}}{M_{\text{dd}}}$$

or more compactly

$$\frac{C_0}{M_0} = E_0^{Q_M}\frac{C_2}{M_2}. \tag{8.4}$$

Here the subscript 0 is attached to the expectation symbol as a reminder that probabilities are used as seen from time 0. As before, the label Q_M serves to indicate that expectation is taken with respect to the implied probabilities that correspond to the numéraire M.

A special two-step tree model that is used often is the one in which the same implied interest rate applies to all nodes. Call this rate r. We then have

$$B_{\text{u}} = B_{\text{d}} = (1+r)B_0$$

$$B_{\text{uu}} = B_{\text{ud}} = B_{\text{du}} = B_{\text{dd}} = (1+r)^2 B_0.$$

In this case, the default choice of the numéraire is the bond. The valuation formula for a contract with payoff C_2 at time 2 becomes

$$C_0 = \frac{1}{(1+r)^2} E_0^Q C_2 = \frac{1}{(1+r)^2}\Big(q_{\mathrm{uu}}^0 C_{\mathrm{uu}} + q_{\mathrm{ud}}^0 C_{\mathrm{ud}} + q_{\mathrm{du}}^0 C_{\mathrm{du}} + q_{\mathrm{dd}}^0 C_{\mathrm{dd}}\Big). \tag{8.5}$$

8.1.2 Hedging in a two-step tree

Now consider the computation of replication weights in a two-step tree. The weights are to be computed at all nodes of the tree except the end nodes. According to the recipe (7.25), the replication weights at the "up" node and at the "down" node are given respectively by

$$x_S^{\mathrm{u}} = \frac{C_{\mathrm{uu}}/B_{\mathrm{uu}} - C_{\mathrm{ud}}/B_{\mathrm{ud}}}{S_{\mathrm{uu}}/B_{\mathrm{uu}} - S_{\mathrm{ud}}/B_{\mathrm{ud}}}, \quad x_B^{\mathrm{u}} = \frac{C_{\mathrm{u}} - x_S^{\mathrm{u}} S_{\mathrm{u}}}{B_{\mathrm{u}}}$$

$$x_S^{\mathrm{d}} = \frac{C_{\mathrm{du}}/B_{\mathrm{du}} - C_{\mathrm{dd}}/B_{\mathrm{dd}}}{S_{\mathrm{du}}/B_{\mathrm{du}} - S_{\mathrm{dd}}/B_{\mathrm{dd}}}, \quad x_B^{\mathrm{d}} = \frac{C_{\mathrm{d}} - x_S^{\mathrm{d}} S_{\mathrm{d}}}{B_{\mathrm{d}}}.$$

At the root node, they are

$$x_S^0 = \frac{C_{\mathrm{u}}/B_{\mathrm{u}} - C_{\mathrm{d}}/B_{\mathrm{d}}}{S_{\mathrm{u}}/B_{\mathrm{u}} - S_{\mathrm{d}}/B_{\mathrm{d}}}, \quad x_B^0 = \frac{C_0 - x_S^0 S_0}{B_0}.$$

Therefore in general *the replication weights are different at different nodes*.

Example 8.1.2 Consider once more the tree in Fig. 8.2. The interest rate is fixed at 3% per period and we are pricing a put option with strike 105. We follow a computational procedure as if the option prices have not been computed previously, so that the option prices can be recomputed from the replication weights and compared to the option prices obtained from the method of state prices, as a check. The replication weights at the "up" node are then found from the equations

$$x_S^{\mathrm{u}} \cdot 120 + x_B^{\mathrm{u}} \cdot 106.09 = 0$$
$$x_S^{\mathrm{d}} \cdot 100 + x_B^{\mathrm{d}} \cdot 106.09 = 5$$

We find $x_S^{\mathrm{u}} = -0.25$, $x_B^{\mathrm{u}} = 0.282779$. From this we can recompute $C_{\mathrm{u}} = x_S^{\mathrm{u}} S_{\mathrm{u}} + x_S^{\mathrm{d}} B_{\mathrm{u}} = 1.6262$. The replication weights at the "down" node are computed in a similar way from the equations

$$x_S^{\mathrm{d}} \cdot 95 + x_B^{\mathrm{d}} \cdot 106.09 = 10$$
$$x_S^{\mathrm{d}} \cdot 85 + x_B^{\mathrm{d}} \cdot 106.09 = 20.$$

We obtain $x_S^{\mathrm{d}} = -1$, $x_B^{\mathrm{d}} = 0.989726$, and $C_{\mathrm{d}} = 11.9417$. At the root of the tree, the relevant equations are

$$x_S^0 \cdot 110 + x_B^0 \cdot 103 = 1.6262$$
$$x_S^0 \cdot 90 + x_B^0 \cdot 103 = 11.9417.$$

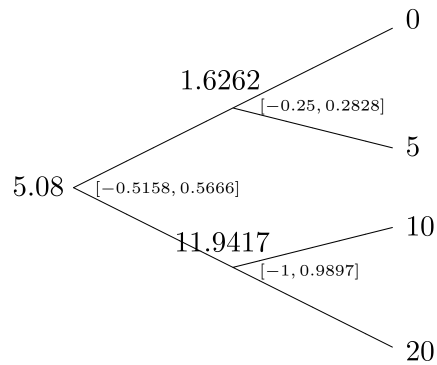

Figure 8.4: Tree with replication weights.

From this we find $x_S^0 = -0.515775$, $x_B^0 = 0.566616$, and $C_0 = 5.08$. The results are shown in Fig. 8.4. If the option values at the nodes are computed first by the method using implied probabilities, then the replication weights can be computed in any order.

The full replication strategy in the two-step tree can now be formulated as follows. We take the point of view of the institution that sells the option. At the root, start with 5.08 (the option premium). Also, go short 0.5158 units of the stock and use the total ($5.08 + 0.5158 \cdot 100 = 56.66$) to buy 0.5666 units of the bond. If an "up" move takes place, the value of the portfolio becomes $-0.5158 \cdot 110 + 0.5666 \cdot 103 = 1.62$. Change the short position in stocks from 0.5158 units to 0.25 units, and change the long position in bonds from 0.5666 units to 0.2828 units (i.e. sell bonds to buy stocks). The value of the new position is $-0.25 \cdot 110 + 0.2828 \cdot 103 = 1.63$. To get it exactly right, we should work in more decimals; the computations are quite sensitive to roundoff errors.

If a "down" move occurs, the value of the portfolio becomes $-0.5158 \cdot 90 + 0.5666 \cdot 103 = 11.94$. Change the short position in stocks from 0.5158 units to one unit, and change the long position in bonds from 0.5666 units to 0.9897 units (i.e. buy bonds and finance it by going further short in stocks). The value of the new position is $-1 \cdot 90 + 0.9897 \cdot 103 = 11.94$.

Suppose we are the "down" node. If a further down move occurs, then the value of the portfolio formed at this node becomes $-1 \cdot 85 + 0.9897 \cdot 106.9 = 20.80$ (again, roundoff errors). If an up move occurs, the portfolio value becomes $-1 \cdot 95 + 0.9897 \cdot 106.9 = 10.80$. At the "up" node: with a further up move, we get $-0.25 \cdot 120 + 0.2828 \cdot 106.9 = 0.23$; after a down move, we have $-0.25 \cdot 100 + 0.2828 \cdot 106.9 = 5.23$.

In conclusion, the portfolio strategy is able (after taking account of rounding errors) to reproduce the option value correctly at every final node, starting from the option premium $C_0 = 5.08$ at the initial node. This is what is called *replication*.

Replication is achieved by a *dynamic* strategy — in other words, a strategy that changes portfolio weights at every new node. The replication strategy is therefore also called *dynamic hedging*. The adjustment of portfolio weights, which in a dynamic strategy takes place at each node, is called *rebalancing*.

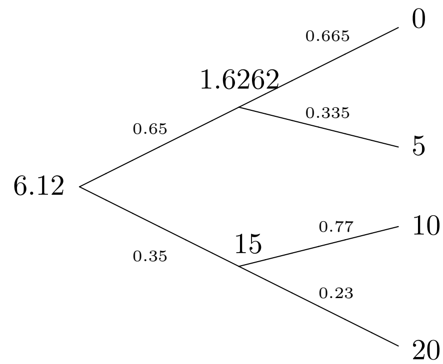

Figure 8.5: American option values in a two-step tree.

8.2 American options

So far we have discussed the pricing of derivatives which can be exercised only at a fixed date. Such derivatives are called *European options*. There are also contracts which allow exercise at any time before or at a given final date; these are called *American options*. Within a two-step binomial tree model, the computation of the value of an American option proceeds in the same way as in the case of a European option. Starting with the given values at time 2, the values at nodes at time 1 are computed using the implied probabilities; from these, the values at time 0 is obtained. The difference is, however, that the computed value at each node (the continuation value) is compared to the value of immediate exercise, and the actual value is made equal to the largest of the two.

Example 8.2.1 Let us consider the same example as before (put option in a two-step tree) but now assume that the put is American rather than European. We have already computed the European put values at the up node and at the down node: 1.6262 and 11.9417, respectively. At the up node, the price of the underlying is 110 so that the value of immediate exercise is 0. At the down node, the price of the underlying is 90 so that the value of immediate exercise is 15. Therefore the American put value at the down node is 15. The continuation value at the initial node is computed as

$$\frac{1}{1.03}\big(0.65 \cdot 1.6262 + 0.35 \cdot 15\big) = 6.12.$$

The value of immediate exercise at the initial node is 5, which is less than the continuation value. The value of the American option is therefore equal to 6.12. The computation is summarized in Fig. 8.5.

8.3 Multistep trees

8.3.1 Pricing in a multistep tree

Pricing of derivatives in multistep binomial trees with two basic assets is not essentially different from pricing in two-step trees, although the notation becomes more elaborate. Each state at timestep N can be represented by a sequence of u's and d's of length N. Such states will be written as $\zeta = \mathrm{v}_1\mathrm{v}_2\cdots\mathrm{v}_N$ where $\mathrm{v}_i \in \{\mathrm{u}, \mathrm{d}\}$ for $i = 1, \ldots, N$. For instance, if $N = 5$, we can have $\zeta =$ duudd, $\zeta =$ uuddu, and so on. For given N, let Z denote the collection of all u-d sequences of length N. The set Z consists of 2^N elements; it corresponds to all *final* nodes of a non-recombining N-step binomial tree. A general node of the tree is described by a sequence of u's and d's of length $\leq N$. The collection of all these sequences, which correspond to all nodes of a non-recombining binomial tree, is denoted by Γ. The set Γ has $2^{N+1} - 1$ elements, counting also the initial node which corresponds to the "empty sequence" (the sequence of length 0). Since every sequence in Γ corresponds to exactly one node in a N-step non-recombining tree, and every node corresponds to exactly one sequence, the terms "node" and "sequence" will be used interchangeably. The two successor nodes of a node $\gamma \in \Gamma$ are γu and γd.

Prices of two basic assets are supposed to be given at each node of the tree. The prices of the two assets at a node γ will be denoted by S_γ and B_γ. It is assumed that the two assets are independent at every node, which means that the two vectors $[S_{\gamma\mathrm{u}} \ \ S_{\gamma\mathrm{d}}]$ and $[B_{\gamma\mathrm{u}} \ \ B_{\gamma\mathrm{d}}]$ are independent for every node γ.

As in the case of two-step trees, we assume that in an investment strategy it is possible to change the portfolio composition at each node without cost, subject only to the constraint that the value of the rearranged portfolio must be equal to the value of the original portfolio. Under this assumption, it can be argued just as in the case of two-step trees that the market is complete, so that in particular the payoff of any derivative product can be replicated, and consequently the price of any derivative product is determined uniquely by the principle of absence of arbitrage. To compute the price, it is not necessary to determine the replicating portfolio first; the method of implied probabilities can be used as an alternative.

To compute implied probabilities, first choose a numéraire. If the model is such that $B_{\gamma\mathrm{u}} = B_{\gamma\mathrm{d}} = (1 + r)B_\gamma$ for some fixed r and for all nodes γ, then the bond B is the default choice for the numéraire. Let the chosen numéraire be indicated by M. The implied probabilities at node γ of an "up" and a "down" move are given implicitly by the following equations (cf. (8.1)):

$$\frac{S_\gamma}{M_\gamma} = q_{\gamma\mathrm{u}} \frac{S_{\gamma\mathrm{u}}}{M_{\gamma\mathrm{u}}} + q_{\gamma\mathrm{d}} \frac{S_{\gamma\mathrm{d}}}{M_{\gamma\mathrm{d}}}\,, \quad \frac{B_\gamma}{M_\gamma} = q_{\gamma\mathrm{u}} \frac{B_{\gamma\mathrm{u}}}{M_{\gamma\mathrm{u}}} + q_{\gamma\mathrm{d}} \frac{B_{\gamma\mathrm{d}}}{M_{\gamma\mathrm{d}}}\,, \qquad q_{\gamma\mathrm{u}} + q_{\gamma\mathrm{d}} = 1. \tag{8.6}$$

Once these implied probabilities have been computed at every node, the price of a derivative asset with given values at step N can be computed by a recursive procedure, going backwards through the tree and applying the formula

$$\frac{C_\gamma}{M_\gamma} = q_{\gamma\mathrm{u}} \frac{C_{\gamma\mathrm{u}}}{M_{\gamma\mathrm{u}}} + q_{\gamma\mathrm{d}} \frac{C_{\gamma\mathrm{d}}}{M_{\gamma\mathrm{d}}} \tag{8.7}$$

at every node. As an alternative to this procedure, one can first compute the probabilities of final states as seen from time 0 on the basis of the stepwise probabilities. For instance if $N = 5$, the implied probability of reaching state uduud from the root of the tree is given by

$$q^0_{\mathrm{uduud}} = q_{\mathrm{u}} q_{\mathrm{ud}} q_{\mathrm{udu}} q_{\mathrm{uduu}} q_{\mathrm{uduud}}.$$

In other words, the implied probability of reaching a particular node from the starting point at time 0 is equal to the product of all stepwise probabilities on the path to that node. With probabilities q^0_ζ defined in this way, the general valuation formula in an N-step binomial tree can be written as

$$\frac{C_0}{M_0} = E^{Q_M} \frac{C_N}{M_N} = \sum_{\zeta \in Z} q^{0M}_\zeta \frac{C_\zeta}{M_\zeta} \tag{8.8}$$

where the superscript M has been added to the q^0's as a reminder that these are implied probabilities associated to the numéraire M. When the bond is taken as a numéraire and interest rates are constant, the formula above simplifies to

$$C_0 = \frac{1}{(1+r)^N} E^Q C_N = \frac{1}{(1+r)^N} \sum_{\zeta \in Z} q^0_\zeta C_\zeta. \tag{8.9}$$

8.3.2 Hedging in a multistep tree

At every node γ, the replication weights x^γ_S and x^γ_B are determined by the equations

$$x^\gamma_S S_{\gamma\mathrm{u}} + x^\gamma_B B_{\gamma\mathrm{u}} = C_{\gamma\mathrm{u}} \tag{8.10a}$$

$$x^\gamma_S S_{\gamma\mathrm{d}} + x^\gamma_B B_{\gamma\mathrm{d}} = C_{\gamma\mathrm{d}}. \tag{8.10b}$$

By construction, we also have

$$x^\gamma_S S_\gamma + x^\gamma_B B_\gamma = C_\gamma. \tag{8.11}$$

In other words, at each node γ, a portfolio can be formed whose value at the node γ is equal to C_γ, and which is composed in such a way that the portfolio value at both successor nodes is equal to the option value at those nodes. If the equality $B_{\gamma\mathrm{u}} = B_{\gamma\mathrm{d}}$ holds, the the easiest way to determine the portfolio weights is to subtract the second equation in (8.10) from the first one, so that a single equation for x^γ_S remains; one finds

$$x^\gamma_S = \frac{C_{\gamma\mathrm{u}} - C_{\gamma\mathrm{d}}}{S_{\gamma\mathrm{u}} - S_{\gamma\mathrm{d}}}. \tag{8.12}$$

The portfolio weight x^γ_B can then be determined from either of the two equations in (8.10), and then (8.11) can be used as a check on the validity of the calculations, or vice versa.

Along each path of “up” and “down” moves that may be realized, the portfolio value remains equal to the option value, while the portfolio composition is changed from node to node depending on the path that is taken. Whichever final node is reached at the time of expiry, the value of the portfolio is still equal to the option value at that node, which implies that at that time the portfolio value is exactly equal to the option payoff. This is the *replication property*.

Figures 8.6, 8.7, 8.8, and 8.9 show the behavior of the hedge strategy in various cases. In each plot, the stock price trajectory has been generated by selecting up and down moves randomly starting from the initial stock price 100; the number of steps is 100, and the stock price is multiplied by 1.02 in the case of an up move and by 0.98 in the case of a down move. The accompanying plots shows the corresponding trajectories of the number of stocks held in the hedging portfolio according to the rule (8.10).

In the case of a call option written on one unit of the stock, the plots show that the number of stocks held in the hedge portfolio is positive; more precisely the number lies between 0 and 1 unit. In the hedge portfolio for a put option written on one unit of the stock, the amount of stocks in the hedge portfolio is negative; it lies between 0 and -1 unit. This means that the hedger takes a short position in the stock. This is reasonable; the put option becomes more valuable when the stock price goes down, and therefore the value of the hedge portfolio should increase as well in that case. Since the hedge portfolio value must go up when the stock price goes down, the number of stocks in the hedge portfolio has to be negative. For the call option, the story is the other way around. The size of the short position in the hedge portfolio for the put option only becomes equal to -1 when the stock value is so low that, within the model, it is certain that the put will end “in the money” (i.e. the stock price at the time of maturity is such that the option value is positive; in the case of a put option this means that the stock value at maturity is below the strike). If it is clear that the option will end “out of the money”, so that the option will have no value at maturity, then the number of stocks in the hedge portfolio is reduced to zero. This happens both in the case of the call and in the case of the stock, but in the case of the put the zero value is reached from below whereas in the case of the call it is reached from above.

If it would be known already at time 0 whether the final value of the stock price would be above or below the strike, then it would be easy to hedge for instance the payoff of the call option. In such a crystal-ball situation, the composition of the perfect hedge portfolio would be *zero* stocks and zero bonds at time 0 if the option ends out of the money, and (in the case of the call option) *one* unit of the stock and $-d(T)K$ worth of the bond if the option is known to end in the money. In reality we do not know at time 0 whether the stock price will be above or below the strike at time T, and so the initial composition of the hedge portfolio is a compromise between the two. As is seen in the figures, the composition is adjusted in the direction of one of the two extremes, in response to stock movements up or down; moreover, the response becomes stronger as the time of maturity approaches.

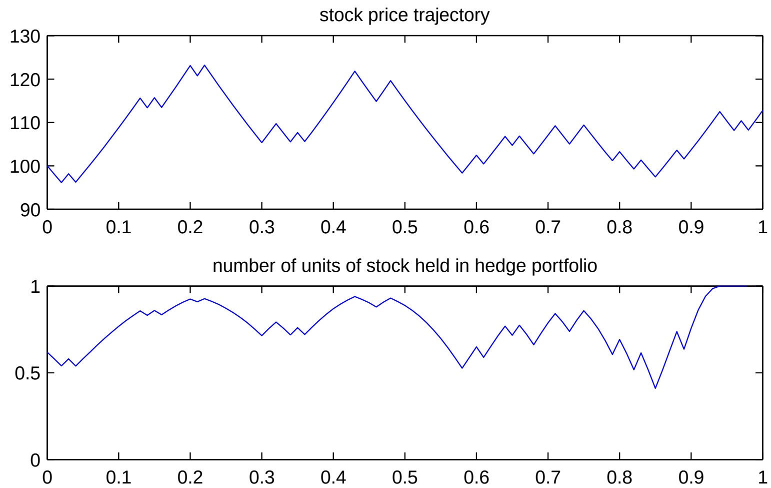

Figure 8.6: Simulation results for hedging of a call option with strike at 100; case in which the option ends in the money.

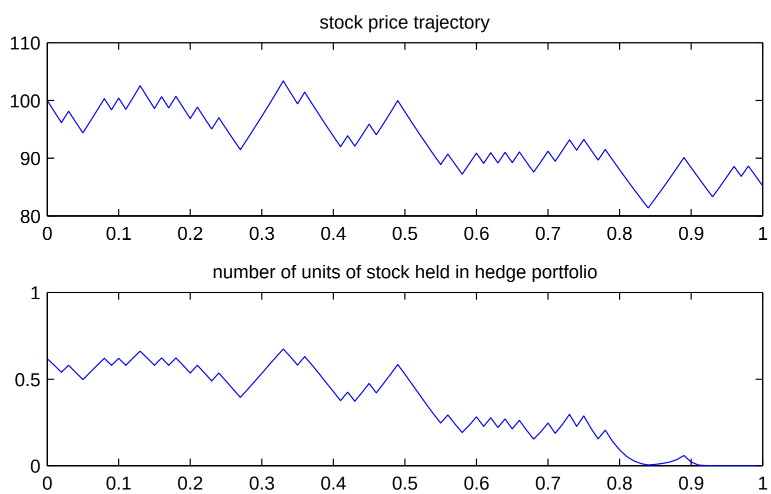

Figure 8.7: Simulation results for hedging of a call option with strike at 100; case in which the option ends out of the money.

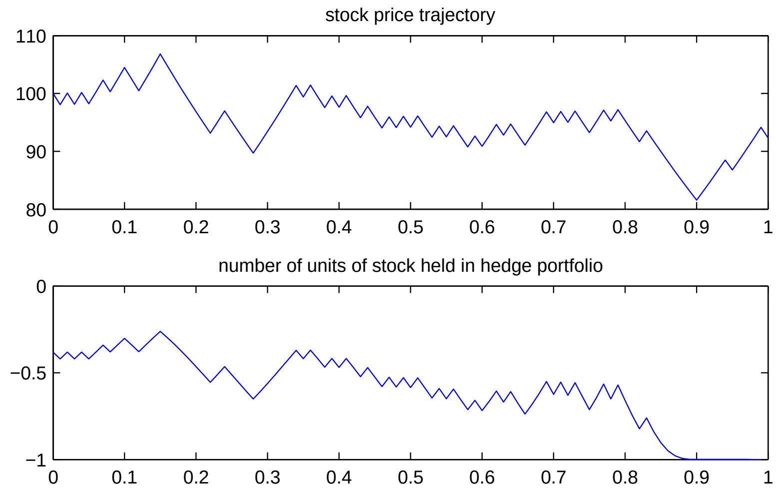

Figure 8.8: Simulation results for hedging of a put option with strike at 100; case in which the option ends in the money.

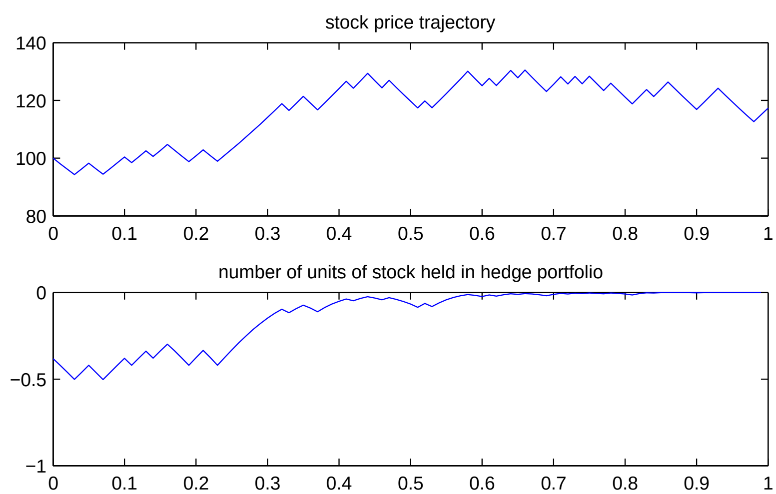

Figure 8.9: Simulation results for hedging of a put option with strike at 100; case in which the option ends out of the money.

8.3.3 The geometric tree model

In the case of binomial trees with many steps, specification of the values of the two basic assets on a node-by-node basis is not very convenient. Instead, some general formula is typically used. The most popular specification is as follows.

Let u and d be positive numbers such that $d < u$, and let S_0 be given. Define the value of S at the nodes of the tree recursively by

$$S_{\gamma\mathrm{u}} = uS_\gamma, \qquad S_{\gamma\mathrm{d}} = dS_\gamma. \tag{8.13}$$

So in an "up" move the stock is multiplied by the number u, and in a "down" move it is multiplied by d. In particular, we have $S_\mathrm{u} = uS_0$, $S_\mathrm{d} = dS_0$, $S_\mathrm{uu} = u^2S_0$, and so on. More generally, one can write

$$S_\gamma = u^{\ell_\mathrm{u}(\gamma)} d^{\ell_\mathrm{d}(\gamma)} S_0. \tag{8.14}$$

where $\ell_\mathrm{u}(\gamma)$ and $\ell_\mathrm{d}(\gamma)$ denote the number of u's in γ and the number of d's in γ respectively. The values of B_γ are specified recursively by

$$B_{\gamma\mathrm{u}} = B_{\gamma\mathrm{d}} = (1+r)B_\gamma$$

where r is a fixed interest rate per step, or in a formula by

$$B_\gamma = (1+r)^{\ell(\gamma)} B_0 \tag{8.15}$$

where $\ell(\gamma) = \ell_\mathrm{u}(\gamma) + \ell_\mathrm{d}(\gamma)$ is the length of the sequence γ. The multi-step binomial tree with the above specification of basic asset prices is called the *geometric tree model*. In other words, the geometric tree model is a binomial model with asset prices specified by (8.14) and (8.15). Note that $S_\mathrm{ud} = S_\mathrm{du} = udS_0$, $S_\mathrm{uud} = S_\mathrm{udu} = S_\mathrm{duu} = u^2dS_0$, and so on; the geometric tree model is recombining.

The implied probabilities for the two branches that lead from γ to the successor nodes γu and γd are computed, when B is taken as the numéraire, from the equations

$$\frac{S_\gamma}{B_\gamma} = q_{\gamma\mathrm{u}}\frac{S_{\gamma\mathrm{u}}}{B_{\gamma\mathrm{u}}} + q_{\gamma\mathrm{d}}\frac{S_{\gamma\mathrm{d}}}{B_{\gamma\mathrm{d}}}, \qquad q_{\gamma\mathrm{u}} + q_{\gamma\mathrm{d}} = 1.$$

In the case of the geometric tree, the first equation can be rewritten as

$$S_\gamma = \frac{q_{\gamma\mathrm{u}}u + q_{\gamma\mathrm{d}}d}{1+r} S_\gamma.$$

It follows that $q_{\gamma\mathrm{u}} = q$ and $q_{\gamma\mathrm{d}} = 1 - q$ where q satisfies

$$qu + (1-q)d = 1 + r. \tag{8.16}$$

In other words, at every node γ we have

$$q_{\gamma\mathrm{u}} = q = \frac{1+r-d}{u-d}, \qquad q_{\gamma\mathrm{d}} = 1 - q = \frac{u-(1+r)}{u-d}. \tag{8.17}$$

Absence of arbitrage holds in the geometric tree model if and only if these numbers are both positive (see Fact 7.2.6, in combination with (7.12) and (7.13)). This means

$$d < 1 + r < u.$$

The cumulative implied probabilities of the final states are given by

$$q_\zeta^0 = q^{\ell_\mathrm{u}(\zeta)}(1-q)^{\ell_\mathrm{d}(\zeta)}.$$

The corresponding values of S_ζ are $u^{\ell_\mathrm{u}(\zeta)}d^{\ell_\mathrm{d}(\zeta)}S_0$ (cf. (8.14)). Because the geometric tree is recombining, the values S_ζ are not all different for different sequences ζ; more specifically, if $\ell_\mathrm{u}(\zeta_1) = \ell_\mathrm{u}(\zeta_2)$ then $\ell_\mathrm{d}(\zeta_1) = \ell_\mathrm{d}(\zeta_2)$ since $\ell_\mathrm{u}(\zeta_1) + \ell_\mathrm{d}(\zeta_1) = N = \ell_\mathrm{u}(\zeta_2) + \ell_\mathrm{d}(\zeta_2)$, and therefore $S_{\zeta_1} = S_{\zeta_2}$. For j with $0 \le j \le N$, there are $\binom{N}{j}$ sequences in Z such that $\ell_\mathrm{u}(\zeta) = j$, and the implied probability of each of these paths is $q^j(1-q)^{N-j}$. Therefore, the implied probability as seen from time 0 to reach a final node ζ such that

$$S_\zeta = u^j d^{N-j} S_0$$

is equal to

$$q^j(1-q)^{N-j}\binom{N}{j}.$$

Under the implied probabilities, the stock price at step N can therefore be described as the random variable S_N defined by

$$S_N = u^J d^{N-J} S_0, \qquad J \sim \mathrm{BIN}(q, N), \quad q = \frac{1+r-d}{u-d}\,. \tag{8.18}$$

Suppose that we want to price an asset whose value at step N is expressed in terms of S_N (i.e. a European derivative written on the underlying S). Examples are: a call option, with payoff function $C_N = \max(S_N - K, 0)$; a put option, with payoff function $C_N = \max(K - S_N, 0)$. For generality, write the payoff as $C_N = f(S_N)$ where f can be any function. Specializing the formula (8.9) to the geometric tree model, with the bond as the numéraire, we can write the price of the derivative at time 0 as

$$C_0 = (1+r)^{-N} E^Q f(S_N) \tag{8.19}$$

or more in detail as

$$C_0 = \frac{1}{(1+r)^N}\sum_{j=0}^{N} q^j(1-q)^{N-j}\binom{N}{j} f(u^j d^{N-j} S_0). \tag{8.20}$$

This is the *binomial option pricing formula*.

At a general node γ, the option price can also be described by the formula above, except that the number of steps to the time of maturity of the option is $N-k$ rather than N, where

k is the length of the sequence γ, and the role of the current stock price is played by S_γ instead of S_0. In other words,

$$C_\gamma = \frac{1}{(1+r)^{N-k}} \sum_{j=0}^{N-k} q^j (1-q)^{N-k-j} \binom{N-k}{j} f(u^j d^{N-k-j} S_\gamma). \tag{8.21}$$

The numbers of units of the stock and the bond to be held in the replicating portfolio at node γ are given by (8.10). In the case of the geometric tree, we have in fact (cf. (7.26))

$$x_S^\gamma = \frac{C_{\gamma\mathrm{u}} - C_{\gamma\mathrm{d}}}{S_{\gamma\mathrm{u}} - S_{\gamma\mathrm{d}}} = \frac{C_{\gamma\mathrm{u}} - C_{\gamma\mathrm{d}}}{(u-d)S_\gamma}\,. \tag{8.22a}$$

Consequently, the *value* of the stock holdings in the replicating portfolio at node γ is given by

$$x_S^\gamma S_\gamma = \frac{C_{\gamma\mathrm{u}} - C_{\gamma\mathrm{d}}}{u-d}\,.$$

The number of units of the bond to be held in the replicating portfolio at node γ is given by (8.10) as well. Alternatively, using (8.11), we can write

$$x_B^\gamma = \frac{C_\gamma}{B_\gamma} - \frac{C_{\gamma\mathrm{u}} - C_{\gamma\mathrm{d}}}{(u-d)B_\gamma}\,. \tag{8.22b}$$

8.3.4 The Black-Scholes formula

It is well known that the binomial distribution converges to the normal distribution when N becomes large. This fact can be used to derive the following limit of the formula (8.20):

$$C_0 = \frac{e^{-rT}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} f\big(S_0 \exp((r - \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}\,z)\big)\,dz. \tag{8.23}$$

The limit is obtained by increasing the number of steps in the tree while keeping the horizon length T constant. The quantities u, d, and r must be adjusted to the step size Δt according to a certain rule, so that convergence is indeed achieved. The properties that the adjustment rule has to satisfy are the following (writing now $u_{\Delta t}$, $d_{\Delta t}$, and $r_{\Delta t}$ instead of u, d, and r, to indicate that these quantities depend on Δt):[2]

$$u_{\Delta t} = 1 + \sigma\sqrt{\Delta t} + O(\Delta t) \tag{8.24a}$$

$$d_{\Delta t} = 1 - \sigma\sqrt{\Delta t} + O(\Delta t) \tag{8.24b}$$

$$r_{\Delta t} = r_{\mathrm{an}}\Delta t + O\big((\Delta t)^2\big) \tag{8.24c}$$

while q is adjusted according to (8.17). The quantity r_{an} that appears in (8.24c) is a rate per unit of time (for instance annual) rather than a rate per step as in (8.20). The quantity σ

[2] The notation $f(x) = g(x) + O(x^k)$ $(x \downarrow 0)$ means that the quantity $(f(x) - g(x))/x^k$ remains bounded as x tends to zero from above. Stated less precisely, it means that the functions $f(x)$ and $g(x)$ behave similarly for small values of x, the degree of similarity being measured by the index k. The addition "$(x \downarrow 0)$" is often omitted when the context makes clear that the values of x near zero are of interest.

is known as the *volatility*. Like r_{an}, it relates to a given unit of time rather than to a single step. Details of the limit calculations are worked out in the appendix to this chapter.

The conditions (8.24) can be satisfied in various ways. A scheme that uses simple expressions for $u_{\Delta t}$, $d_{\Delta t}$, and $r_{\Delta t}$ is the following:

$$u_{\Delta t} = 1+\sigma\sqrt{\Delta t}, \quad d_{\Delta t} = 1-\sigma\sqrt{\Delta t}, \quad r_{\Delta t} = r_{\text{an}}\Delta t, \quad q_{\Delta t} = \frac{\sigma\sqrt{\Delta t} + r_{\text{an}}\Delta t}{2\sigma\sqrt{\Delta t}}\,. \tag{8.25}$$

A variation of this which is closer to the "continuous" spirit is:

$$u_{\Delta t} = e^{\sigma\sqrt{\Delta t}}, \quad d_{\Delta t} = e^{-\sigma\sqrt{\Delta t}}, \quad r_{\Delta t} = e^{r_{\text{an}}\Delta t} - 1, \quad q_{\Delta t} = \frac{e^{r_{\text{an}}\Delta t} - e^{-\sigma\sqrt{\Delta t}}}{e^{\sigma\sqrt{\Delta t}} - e^{-\sigma\sqrt{\Delta t}}}\,. \tag{8.26}$$

Finally one can also choose to work with a specification in which the implied probability is simple:

$$u_{\Delta t} = 1 + \sigma\sqrt{\Delta t} + r_{\text{an}}\Delta t, \quad d_{\Delta t} = 1 - \sigma\sqrt{\Delta t} + r_{\text{an}}\Delta t, \quad r_{\Delta t} = r_{\text{an}}\Delta t, \quad q_{\Delta t} = \tfrac{1}{2}\,. \tag{8.27}$$

The formula (8.23) is known as the *lognormal pricing formula*. Compared to the binomial pricing formula, it is more convenient from a practical point of view, as well as closer to reality in the sense that the parameters appearing in the formula are more directly linked to real-world quantities. In particular, σ is just the standard deviation of the percentage return across one unit of time (for instance a year). Special cases of the lognormal pricing formula can be derived from specific forms of the option payoff function $f(S_T)$. In particular, if f is the payoff function of a call option, $f(S_T) = \max(S_T - K, 0)$ where K is a constant, then it can be computed (see the appendix to this chapter) that the formula (8.23) becomes

$$c_0 = S_0\Phi(d_1) - e^{-rT}K\Phi(d_2) \tag{8.28a}$$

where Φ refers to the cumulative standard normal distribution function, r denotes the annual interest rate, and d_1 and d_2 are constants defined by

$$d_1 = d_2 + \sigma\sqrt{T} = \frac{\ln(S_0/K) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \tag{8.28b}$$

$$d_2 = \frac{\ln(S_0/K) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\,. \tag{8.28c}$$

The expression (8.28) is the *Black-Scholes formula* for the price of a call option. The publication of this formula in 1973 is generally considered to mark the birth of modern mathematical finance.[3]

[3]For historical correctness, it should be noted that neither was the event fully appreciated at the time, nor did Black and Scholes derive their formula in the way suggested here.

8.4 An application to multiperiod optimization

The formulas that have been used above to derive hedging strategies may be used also in multiperiod optimization. The following gives an illustration of the idea. Assume, as usual in this chapter, that there are two basic assets denoted by S and B, which are independent at every node, and that portfolio rebalancing is possible at every node without costs, subject only to the budget constraint (portfolio value after rebalancing must be the same as portfolio value before rebalancing). We then have a complete market. For simplicity, it will also be assumed that the same interest rate r applies at every step, i.e. $B_{\gamma\mathrm{u}} = B_{\gamma\mathrm{d}} = (1+r)B_\gamma$ at every node γ.

Consider, as in the single-period case of Ch. 2, a final wealth problem in terms of expected utility with respect to a concave utility function $u(x)$, expectation being taken with respect to the subjective probabilities of a given investor. The optimization procedure now works as follows. First compute the implied probabilities of the final nodes as seen from time 0. Let these probabilities be denoted by q^0_ζ. Then optimize the values of the outcomes V_ζ at the final nodes of the tree, subject to the budget constraint

$$\sum_{\zeta\in Z} q^0_\zeta V_\zeta = (1+r)^N V_0.$$

The optimal solution is given by

$$V_\zeta = v(\lambda q^0_\zeta / p^0_\zeta)$$

where the p^0_ζ's are the subjective probabilities of the final states as seen from time 0, $v(y)$ is the inverse function of the marginal utility $u'(x)$, and λ is determined by V_0 through the budget constraint.

The optimal payoff $(V_\zeta)_{\zeta\in Z}$ is analogous to an option payoff that we want to replicate. It is guaranteed that replication is possible, because the budget constraint is satisfied and the market is complete. Just as we did in the case of options, the portfolio holdings that need to be chosen at each node can be determined in a backward procedure. The holdings are in general different at different nodes, causing the portfolio to be rebalanced at every step. In this way an optimal investment strategy is defined.

8.5 What is a good model?

Just to show that the geometric tree model is not the only possible one, let us also consider the *arithmetic tree model*. This model is defined by the following specification of asset prices:

$$\begin{aligned} S_\gamma &= S_0 + \ell_{\mathrm{u}}(\gamma)a - \ell_{\mathrm{d}}(\gamma)b \\ B_\gamma &= (1+r)^{\ell(\gamma)} B_0 \end{aligned}$$

where a and b are given positive numbers. Whereas in the geometric tree model the stock price after an "up" or "down" move is *multiplied* by a fixed number, in the arithmetic tree model a fixed amount is *added* or subtracted.

The arithmetic tree model is much less popular than the geometric tree. Its main disadvantages are:

- the implied probabilities for "up" and "down" moves are different at different nodes, whereas in the geometric tree they are the same at all nodes
- the total number of steps in the tree, N, has to be less than S_0/b, or otherwise the stock price can become negative
- also, N has to be smaller than $\frac{1}{r} + 1 - S_0/a$, or otherwise there will be arbitrage opportunities at the high end of the tree.

Still, for a limited number of steps the arithmetic tree model can be used. Compared to the geometric tree model, it then leads to a *different* option price and also to a *different* hedging strategy.

So we can choose between the geometric model, the arithmetic model, and other models that might be proposed. Even if we opt for the geometric tree, we still need to select the parameters u, d, and r. The following are the most important guidelines for selecting a model that does well in practice:

- the option price that follows from the model should be competitive
- the model should produce a hedging strategy that generates a small variance of the *hedge error* (the difference between the value of the hedge portfolio and the value of the option at the time of expiry).

The second criterion is more difficult to assess than the first one. Still the behavior of the hedge produced by the model can be tested on past data, or on a collection of future scenarios which are deemed realistic.

As noted above, even if one chooses the geometric model, there is still the question of how to choose the parameters of the model. The following prescription is often used (cf. (8.25).

Rule 8.5.1 In the geometric tree model defined by (8.14) and (8.15), take r equal to the discretely compounded interest rate that corresponds to one time step of the model, take u equal to one plus the standard deviation of the relative change of the stock price during one time step, and take d equal to one minus this standard deviation.

For instance, if the time step is one day and the daily standard deviation of the relative change of the stock price is 2%, then take $u = 1.02$, $d = 0.98$. The value of r should be the one-day interest rate; for instance if the annual rate is 4%, then the daily rate is

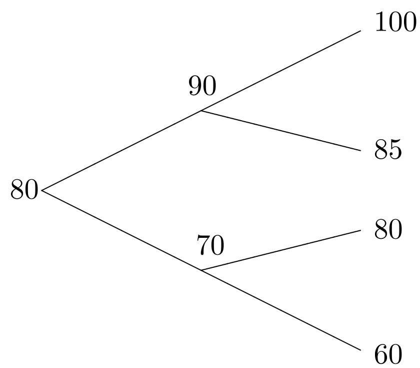

Figure 8.10: Two-period market for Exc. 1.

approximately one basis point (i.e. $r \approx 0.0001$). The factor $(1+r)^{-N}$ that appears in the option pricing formula (8.20) can be calculated in a simpler way as the discount factor that belongs to the N-th step of the tree, which corresponds to the maturity date of the option.

The standard deviation of the relative change of the stock price is usually called the *volatility* of the stock price. If it is assumed that the daily relative changes of the stock price follow a normal distribution with a mean that is approximately zero, then to say that the volatility on a daily basis is 2% means that on 68% of trading days the relative change of the stock price with respect to the previous day is less than 2% (up or down). Stock markets tend to go through periods of higher and lower volatility; the level of 2% on a daily basis corresponds, historically speaking, to fairly nervous times.

The rule 8.5.1 is good if it meets the criteria in the bullet points above. In particular, the price computed from (8.20) should be close to the price at which the option is actually traded in the market, and the hedge strategy given by (8.21) and (8.22) should be effective when applied to actual stock and bond prices. The implementation of the hedge moreover has to deal with the fact that the relative changes of the stock price are not limited to only two values as in the tree model. Whether or not the hedge will be effective depends on properties of the real world. If certain properties are assumed concerning the way that actual stock and bond prices move, a mathematical analysis can be undertaken to prove that the rule 8.5.1 is indeed effective.

8.6 Exercises

1. Consider the tree model for the price of a risky asset shown in Fig. 8.10. Assume that the discretely compounded riskfree interest rate is 4% per period. One period corresponds to one step in the tree.

a. Determine the price of a European put option with strike $K = 90$.

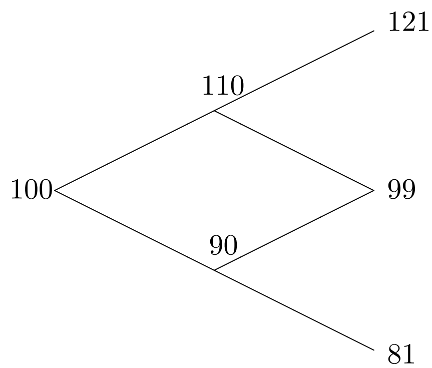

Figure 8.11: Two-period market for Exc. 2.

b. Determine the price of an American put option with the same strike.

2. Consider the two-step recombining tree model in Fig. 8.11. Assume (for simplicity) that the interest rate is zero.

a. Alice sells to Bob a European call option with strike $K = 95$. Determine the arbitrage-free price of the option according to the above model.

b. If Alice doesn't use the option premium that she receives from Bob to set up a hedge portfolio, what would be her maximum potential loss as a result of the contract?

c. Determine the replicating portfolio strategy for the call option.

d. Suppose that Alice forms her hedge portfolio at time 0 as in the question above, but that she neglects to rebalance the portfolio at time 1. What is now her maximum loss?

3. Consider again the tree model of the previous exercise, but now suppose that the interest rate is 3% per step, and that Alice sells an American put option with strike 105 to Bob. The price of the bond at time 0 is 100.

a. Determine the premium for the put option.

b. Determine the hedge strategy that Alice should apply to guarantee that in no case she will have losses from the contract.

c. Suppose that Bob doesn't make use of the early exercise opportunities that come with an American option. Which advantage does that bring to Alice?

d. Suppose that, in the situation in which he should exercise early, Bob actually likes the payoffs that would be produced by the put option if he would not exercise it. What would then be the best thing for him to do?

4. Consider a European option defined by its payoff function $C_T = f(S_T)$. Compute the option price according to the binomial pricing formula (8.20) in the following cases: (i)

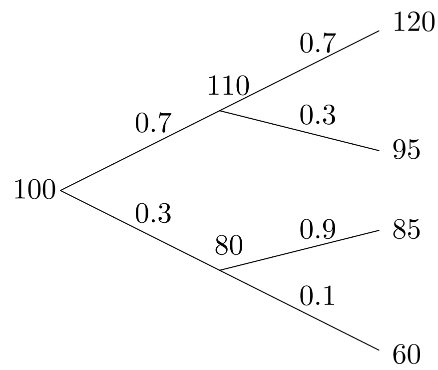

Figure 8.12: Two-period market for Exc. 6.

$f(S_T) = 1$, (ii) $f(S_T) = S_T$, (iii) $f(S_T) = S_T - K$ where K is a given constant. (Hint: recall the binomial formula $(a+b)^N = \sum_{k=0}^{N} \binom{N}{k} a^k b^{N-k}$.) Do you get the answers that you expect?

5. A European *digital option* with strike K and expiration date T is an option that pays one unit of currency if the price of the underlying at time T is above K, and nothing otherwise. Determine the price of the digital option at time 0 according to the binomial pricing formula. Write your answer in terms of the binomial cumulative distribution function $B(x; n, p)$.

6. Suppose that the evolution of the price of a risky asset is given by the tree model of Fig. 8.12, where the probabilities indicated are *objective* ("real-world") probabilities. The riskfree interest rate is assumed to be zero. An investor with an initial capital $V_0 = 100\,000$ euro is aiming to maximize $E[\ln(V_2)]$, where V_2 indicates portfolio value at time 2, by following a portfolio strategy which makes use of the risky asset of Fig. 8.12 ("stock") and the riskless asset which consists of the savings account with 0% interest rate ("bond").

a. Motivate the following statement: the outcomes of capital at time 2 that achieve the investor's objective can be obtained by solving the optimization problem

$$\begin{aligned} &\text{maximize} \quad p^0_{\text{uu}} \ln(V_{\text{uu}}) + p^0_{\text{ud}} \ln(V_{\text{ud}}) + p^0_{\text{du}} \ln(V_{\text{du}}) + p^0_{\text{dd}} \ln(V_{\text{dd}}) \\ &\text{subject to} \quad q^0_{\text{uu}} V_{\text{uu}} + q^0_{\text{ud}} V_{\text{ud}} + q^0_{\text{du}} V_{\text{du}} + q^0_{\text{dd}} V_{\text{dd}} = (1+r)^2 V_0 \end{aligned}$$

where the numbers p^0_{uu}, p^0_{ud}, p^0_{du}, and p^0_{dd} are the objective probabilities of the four final states as seen from time 0, the numbers q^0_{uu}, q^0_{ud}, q^0_{du}, and q^0_{dd} are the implied probabilities of the four final states also as seen from time 0, and $r = 0$.

b. Determine the implied probabilities q^0_{uu}, q^0_{ud}, q^0_{du}, and q^0_{dd}, and solve the optimization problem above using the Lagrangian method.

c. Derive the optimal levels of capital at nodes u and d.

d. Determine how much of the available capital should be invested in stocks at time 0, and how much should be invested in bonds.

e. Also determine the optimal portfolio composition at nodes u and d.

7. Consider a market that is given by a geometric tree model with parameters u, d, and r; the initial values for the two assets S and B are $S_0 > 0$ and $B_0 > 0$ respectively. It is assumed that $u > 1 + r > d > 0$. The number q is defined by $q = (1 + r - d)/(u - d)$.

a. Let p be a given number such that $0 < p < 1$. Prove that it is possible, given an initial capital V_0, to define a trading strategy which satisfies the budget constraint at every node and which is such that the portfolio value at a node γ that is reached after j "up" moves and $k - j$ "down" moves is given by

$$V_\gamma = \frac{p^j(1-p)^{k-j}}{q^j(1-q)^{k-j}}\,(1+r)^k V_0.$$

Derive explicit expressions for the amounts of units of both assets that are to be held in the portfolio at each node, and prove that the percentage of portfolio value held in risky assets (S) is the same at each node.

b. Prove that the portfolio strategy defined above provides an optimal solution for the maximization problem

$$\begin{array}{rl} \text{maximize} & E^P[\ln(V_N)] \\ \text{subject to} & \text{budget constraint at each node,} \\ & \text{initial portfolio value} = V_0 \end{array}$$

where $N > 0$ is a given integer, and the symbol E^P indicates that expectation is taken under the assumption that the objective probability of an "up" move at each node is given by p.

8.7 Appendix: limit calculations

The implied distribution of the stock price after N time steps in the geometric tree model, as seen from time 0, is given by (8.18). We now want to let $N \to \infty$ and $\Delta t = T/N \to 0$. The stochastic variable S_N represents the stock price after N steps of length Δt, with the distribution as determined by the implied probabilities; in other words, this stochastic variable corresponds to the stock price at time T. Since T is kept fixed below while N is variable, we will from now on write S_T rather than S_N. Likewise we write B_T rather than B_N. To alleviate the notation, we also write from now on r instead of r_{an}.

First consider the limit of B_T as T tends to infinity. The equation $B_T = (1 + r_{\Delta t})^N B_0$ can be rewritten as

$$\ln B_T = \ln B_0 + N \ln(1 + r_{\Delta t}). \tag{8.29}$$

Making use of the expansion

$$\ln(1+x) = x + O(x^2) \tag{8.30}$$

we can write, in view of (8.24c) and the fact that $N = T/\Delta t$,

$$\ln B_T = \ln B_0 + \frac{T}{\Delta t}\big(r\Delta t + O((\Delta t)^2)\big) = \ln B_0 + rT + O(\Delta t). \tag{8.31}$$

It follows that

$$\lim_{\Delta t \to 0} \ln B_T = \ln B_0 + rT.$$

This in turn implies

$$\lim_{\Delta t \to 0} B_T = e^{rT} B_0. \tag{8.32}$$

The parameter r should therefore be interpreted as the continuously compounded annual interest rate.

When we discuss what happens to S_T as N tends to infinity (or equivalently, as Δt tends to zero), it has to be taken into account that S_T is a stochastic variable. We will therefore be looking for the limit *distribution* of S_T. Start by defining

$$Z_N := \sqrt{N}\, \frac{J/N - q}{\sqrt{q(1-q)}}\,.$$

According to the central limit theorem, the distribution of Z_N tends to the standard normal distribution as N tends to infinity.[4] On the basis of this definition we can write

$$\frac{J}{N} = q + \sqrt{\frac{q(1-q)}{N}}\, Z_N.$$

From $S_T = u^J d^{N-J} S_0$ we obtain

$$\begin{aligned} \ln S_T &= J \ln u + (N-J)\ln d + \ln S_0 \\ &= N\Big(\frac{J}{N}\ln u + \big(1 - \frac{J}{N}\big)\ln d\Big) + \ln S_0. \end{aligned}$$

Therefore

$$\begin{aligned} \ln S_T &= \frac{T}{\Delta t}\left(\Big(q + \sqrt{\frac{q(1-q)}{T/\Delta t}}\, Z_N\Big)\ln u + \Big(1 - q - \sqrt{\frac{q(1-q)}{T/\Delta t}}\, Z_N\Big)\ln d\right) + \ln S_0 \\ &= \frac{q\ln u + (1-q)\ln d}{\Delta t}\, T + \frac{\sqrt{q(1-q)}(\ln u - \ln d)}{\sqrt{\Delta t}}\sqrt{T}\, Z_N + \ln S_0. \end{aligned} \tag{8.33}$$

[4] Actually this requires a somewhat stronger version of the central limit theorem than the standard one, since the parameter q depends on N, except when the specification (8.27) is used.

The problem comes down to: given

$$u = u_{\Delta t} = 1 + \sigma\sqrt{\Delta t} + O(\Delta t) \tag{8.34a}$$

$$d = d_{\Delta t} = 1 - \sigma\sqrt{\Delta t} + O(\Delta t) \tag{8.34b}$$

$$r_{\Delta t} = r\Delta t + O(\Delta t^2) \tag{8.34c}$$

$$q = q_{\Delta t} = \frac{1 + r_{\Delta t} - d_{\Delta t}}{u_{\Delta t} - d_{\Delta t}} \tag{8.34d}$$

find the limit values

$$\lim_{\Delta t \downarrow 0} \frac{q \ln u + (1-q)\ln d}{\Delta t} \tag{8.35}$$

and

$$\lim_{\Delta t \downarrow 0} \frac{\sqrt{q(1-q)}(\ln u - \ln d)}{\sqrt{\Delta t}}\,. \tag{8.36}$$

These limits can be determined by systematic use of "big O" calculus. We start with the second limit because it is a bit simpler than the first one.

The implied probability q is determined by the expression (8.34d). In view of the fact that

$$u - d = 2\sigma\sqrt{\Delta t} + O(\Delta t)$$

we have

$$q = \frac{1 + r_{\Delta t} - d}{u - d} = \frac{\sigma\sqrt{\Delta t} + O(\Delta t)}{2\sigma\sqrt{\Delta t} + O(\Delta t)} = \frac{1}{2}\cdot\frac{1 + O(\sqrt{\Delta t})}{1 + O(\sqrt{\Delta t})} = \tfrac{1}{2} + O(\sqrt{\Delta t}). \tag{8.37}$$

Note that an expression of the form $(1+O(x))/(1+O(x))$ is not always equal to 1 because the functions represented by the O symbols in the numerator and the denominator may not be the same. However, since

$$\frac{1}{1+x} = 1 + O(x)$$

we have $1/(1+O(x)) = 1 + O(x)$ and therefore $(1+O(x))/(1+O(x)) = 1 + O(x)$. From $q = \frac{1}{2} + O(\sqrt{\Delta t})$ it follows that $1 - q = \frac{1}{2} + O(\sqrt{\Delta t})$ and consequently

$$q(1-q) = \tfrac{1}{4} + O(\sqrt{\Delta t}) = \tfrac{1}{4}\big(1 + O(\sqrt{\Delta t})\big).$$

Since

$$\sqrt{1+x} = 1 + O(x)$$

we find

$$\sqrt{q(1-q)} = \tfrac{1}{2}(1 + O(\sqrt{\Delta t})) = \tfrac{1}{2} + O(\sqrt{\Delta t}).$$

Next we need to find an expression for $\ln u - \ln d$. From (8.30) it follows that

$$\ln u = \ln(1 + (u-1)) = u - 1 + O((u-1)^2) = \sigma\sqrt{\Delta t} + O(\Delta t).$$

In the same way, one obtains $\ln d = -\sigma\sqrt{\Delta t} + O(\Delta t)$. Therefore

$$\ln u - \ln d = 2\sigma\sqrt{\Delta t} + O(\Delta t).$$

Putting things together, we find

$$\sqrt{q(1-q)}(\ln u - \ln d) = \sigma\sqrt{\Delta t} + O(\Delta t).$$

Therefore

$$\lim_{\Delta t \downarrow 0} \frac{\sqrt{q(1-q)}(\ln u - \ln d)}{\sqrt{\Delta t}} = \sigma.$$

This establishes the value of the limit in (8.36).

The limit in (8.35) has Δt in the denominator and so we need to work out higher-order terms in the numerator as well. We use again the Taylor series expansion of $\ln(1+x)$, but take it one step further:

$$\ln(1+x) = x - \tfrac{1}{2}x^2 + O(x^3).$$

From this it follows that

$$\ln u = u - 1 - \tfrac{1}{2}(u-1)^2 + O((\Delta t)^{3/2}) = u - 1 - \tfrac{1}{2}\sigma^2\Delta t + O((\Delta t)^{3/2}).$$

Analogously we have

$$\ln d = d - 1 - \tfrac{1}{2}(d-1)^2 + O((\Delta t)^{3/2}) = d - 1 - \tfrac{1}{2}\sigma^2\Delta t + O((\Delta t)^{3/2}).$$

We also use

$$q(u-1) + (1-q)(d-1) = qu + (1-q)d - 1 = r_{\Delta t} = r\Delta t + O(\Delta t^2).$$

From the three equations above one finds

$$q\ln u + (1-q)\ln d = (r - \tfrac{1}{2}\sigma^2)\Delta t + O((\Delta t)^{3/2}).$$

Therefore:

$$\lim_{\Delta t \downarrow 0} \frac{q\ln u + (1-q)\ln d}{\Delta t} = r - \tfrac{1}{2}\sigma^2.$$

Having completed these calculations, we can now look at (8.33) again. We have found that asymptotically the distribution of $\ln S_T$ is given by

$$\ln S_T = (r - \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}\,Z + \ln S_0\,, \quad Z \sim N(0,1).$$

In other words, the limit distribution of $\ln S_T$ is normal:

$$\ln S_T \sim N(\ln S_0 + (r - \tfrac{1}{2}\sigma^2)T, \sigma^2 T).$$

The asymptotic distribution of S_T is consequently given by

$$S_T = S_0 \exp\big((r - \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}\,Z\big), \quad Z \sim N(0,1). \tag{8.38}$$

This is a *lognormal* distribution. The calculation above has shown that the same asymptotic distribution is obtained for any specification of the functions $u_{\Delta t}$, $d_{\Delta t}$, and $r_{\Delta t}$, as long as the conditions (8.24) are satisfied.

Consider now a European call option with strike K, maturing at time T. Let the price of this option at time 0 be denoted by c_0. Given the fact that the value of the option at time T is given by $\max(S_T - K, 0)$, the price of the option according to the lognormal pricing formula (8.23) is

$$c_0 = \frac{e^{-rT}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} \max\left(S_0 \exp((r - \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}\,z) - K, 0\right) dz.$$

This integral can be worked out exactly. The solution makes use of the cumulative normal distribution function

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}z^2}\, dz.$$

The first thing to note is that the integrand is zero for $S_0 \exp[(r - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}\,z] \le K$, or in other words for $z \le -d_2$ where $d_2 := \big(\ln(S_0/K) + (r - \frac{1}{2}\sigma^2)T\big)/(\sigma\sqrt{T})$. So we have

$$c_0 = \frac{e^{-rT}}{\sqrt{2\pi}} \int_{-d_2}^{\infty} e^{-\frac{1}{2}z^2} \left(S_0 \exp[(r - \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}\,z] - K\right) dz.$$

This can be rewritten as

$$c_0 = \frac{e^{-\frac{1}{2}\sigma^2 T} S_0}{\sqrt{2\pi}} \int_{-d_2}^{\infty} e^{-\frac{1}{2}z^2} e^{\sigma\sqrt{T}\,z}\, dz - \frac{e^{-rT}K}{\sqrt{2\pi}} \int_{-d_2}^{\infty} e^{-\frac{1}{2}z^2}\, dz.$$

The second term on the right can be rewritten directly in terms of the cumulative normal distribution function:

$$\frac{e^{-rT}K}{\sqrt{2\pi}} \int_{-d_2}^{\infty} e^{-\frac{1}{2}z^2}\, dz = e^{-rT}K(1 - \Phi(-d_2)) = e^{-rT}K\Phi(d_2).$$

The first term involves an integral of the form

$$\frac{e^{-\frac{1}{2}a^2}}{\sqrt{2\pi}} \int_{-d_2}^{\infty} e^{-\frac{1}{2}z^2} e^{az}\, dz.$$

Note that $e^{-\frac{1}{2}a^2} e^{-\frac{1}{2}z^2} e^{az} = e^{-\frac{1}{2}(z-a)^2}$. Therefore,

$$\frac{e^{-\frac{1}{2}a^2}}{\sqrt{2\pi}} \int_{-d_2}^{\infty} e^{-\frac{1}{2}z^2} e^{az}\, dz = \frac{1}{\sqrt{2\pi}} \int_{-d_2}^{\infty} e^{-\frac{1}{2}(z-a)^2}\, dz \overset{y\,=\,z\,-\,a}{=}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-(d_2+a)}^{\infty} e^{-\frac{1}{2}y^2}\, dy = \Phi(d_2 + a).$$

The lognormal pricing formula in the case of a call option therefore becomes

$$c_0 = S_0\Phi(d_1) - e^{-rT}K\Phi(d_2) \tag{8.39a}$$

where

$$d_1 = d_2 + \sigma\sqrt{T} = \frac{\ln(S_0/K) + (r + \frac{1}{2}\sigma^2 T)}{\sigma\sqrt{T}} \tag{8.39b}$$

$$d_2 = \frac{\ln(S_0/K) + (r - \frac{1}{2}\sigma^2 T)}{\sigma\sqrt{T}}. \tag{8.39c}$$

This is the famous *Black-Scholes formula* (published 1973; Nobel prize 1997). The derivation in the original paper is different from the one shown here.