Financial Innovation, Structuring and Risk Transfer

Vanini, Paolo

University of Basel

November 2012

Financial Innovation, Structuring and Risk Transfer

Paolo Vanini
Preliminary Notes
4th Revision of Chapter 1.

November 7, 2012

Contents

Chapter 1

Overview

1.1 Market Structure and Efficiency

What is financial innovation? Although a large financial industry emerged in the last decades claiming to be innovative there is no simple answer what financial innovation is. Broadly financial innovation is often related to incompleteness and/or inefficiencies in financial markets. Consider the extreme case where financial markets are complete and perfect. Financial innovation then faces a Modigliani-Miller theorem type situation: Innovation adds no value since in a complete market wealth can be transferred between any desirable states of the world at a given date and also between states at different dates to generate **any** desirable payoff at no costs (perfect markets).[1] Whatever payoff at a given date an agent wishes, he can construct this payoff at zero costs using market instruments.

We consider two incomplete markets: Home financing risk and demographic risk markets.

Private individuals use **mortgages** to finance their houses. This often leads to a high asset (house price) to equity ratio, i.e. a high leverage. High leverage means small equity buffers to avoid insolvency, i.e. the asset value exceeds the liability value (the mortgage in our case) only by a small amount. Can homeowners hedge their default risk? Although the value of real estate is in most nations a multiple of the value of all listed stocks it is almost impossible to trade pure real estate market risk: How can you protect your balance sheet against a drop of the price of your house? There is no financial contract (say a put option) which provides protection against specific real estate risk. Figure 1.1 shows a complete market case where one could protect against falling house prices. The put option payoff at maturity is $\max(K - S_T, 0)$ with K the strike price and S_T the price of the house. Fair pricing using no arbitrage then gives the option price today - given the final payoff it must be proportional to a functional of the difference $K - S_0$[2]. For the issuer of the option, the bank, this means that the trader has to be short the amount S_0. But how can you short a house price? This defines the supply side reason

[1] No transaction costs (bid-ask spreads), tax costs, information costs.

[2] The functional is the linear one. This is the result of the linear pricing theorem.

for an incomplete market. Some proxy instruments for risk transfer exist but their hedge

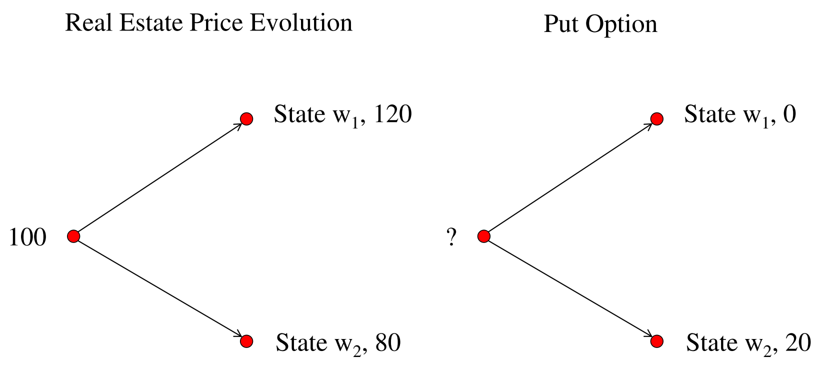


Figure 1.1: One-period model of house prices. The house price S today of 100 can either rise or fall in one period. The protection is obtained by the put option, i.e. the option which pays in this case $\max(K - S, 0)$ with $K = 100$ the strike price. The option never has a negative value since the buyer of the option has the right but not the obligation to exercise the put. If house prices increase, the option is not exercised and it is worthless in this state. If prices drop, the option pays $20 = 100 - 80$. This defines the protection. How much does one have to pay for such a product today? That for, one has to develop a pricing theory. This theory is based on the no arbitrage principle and mathematical logic. We discuss this in details below. The real estate market is not complete in the sense, that no option contract exists for the two states, i.e. the price process of real estate exists but no option process.

effectiveness is (very) limited. Why is there such a immature financial market for such an important asset class? An asset class which is important both for the micro- and the macro economy and the financial market stability. First, the construction of transparent and trustful market instruments is not easy. This is due to the heterogeneity of the goods - each house is unique. Today there exist solutions to this problem using hedonic price indices. We consider this technical issue in Section 6.1. Second, many people fail to have a reasonable estimate about the risk of mortgage financing. This has several reasons. To buy a house is often an emotional event and people often spent time to find an appropriate object. The inhabitants then want to equip the house and live their. They do not want to consider mentally the case that the object could lose value. Since houses are physical goods social peer-group pressure is a further reason why risks are overlooked or underestimated. We consider demand side reasons for market incompleteness. Interest rate risk is often overestimated and real estate market risk is underestimated. That is a drop in the house price of 10 percent has due to the leverage for many individuals a far more disastrous effect on their equity than a rise of interest rates by 1 percent. Why do people often show this bias in risk identification? While interest risk is exogenous, i.e. people cannot affect it, market risk of the house is endogenous: The decision to buy **that** house depends on the preferences of the individuals. Furthermore, interest rate risk has typically less severe impacts on homeowners than market risk has. People have more possibilities to withstand increasing interest rate risks than to falling house prices. Again, different leverage levels matter. But there is also a higher willingness on the lender's side to endure a period of high interest rate risk. Simply because the probability is larger that the homeowner will recover in this risk scenario than in the case with falling house prices.

Demographic change. Demographic change is likely to turn some existing risk sharing mechanism for retirement provision, such as pay-as-you-go systems, down. It will be difficult to enforce the system in the future if the number of active workers decreases more and more compared to the retired population. This possible evolution raises several risks which so far cannot be structured, transferred or covered by financial contracts. This shift in demography has several impacts on the stability of the retirement system. Often risks in the first and second pillars of western social security systems are not priced at all or mispriced. An example are guarantees which one generation provides to another generation. Such guarantees are option contracts: Their value depends on an underlying asset value and the payoff profile is not linear - a guarantee by definition cuts at some point the underlying asset value. But for the seller of such non linear contracts the pricing and hedge requires options: The price of a guarantee in social security contract is the price of a derivative or more precisely an option. But the options between the issuer of the contract - say a today young generation - and the buyer - the today retired generation - is not done. The pricing of these options is mostly not done using financial principles but solidarity between the generations is implicitly used. This has two immediate effects for the society: First, there is intransparency about the price of social security risk premia. Second, the premia are not allocated but exist in a diffusive way in the retirement system. What many inhabitants fear is that due to the changing demography and the still increasing life expectancy the hidden price of solidarity between generations growth and will break due to no longer acceptable transfer payments between the generations. Consider pension funds. There are basically two types of pension fund plans. Defined benefit, where the employer promises future benefits and he invests to meet the promise. In this case the firm bears capital market risk. In defined contribution plans only the contribution but not the benefit is assured. The employee bears the capital market risk. Figure 1.2 illustrate the present values of the two types of plans as a function of the pension fund asset value.

The profile of a defined benefit scheme is the same as the payoff profile of a bond. If interest rates are low the employer faces a problem that the payoff of the investment is smaller than the promised payoff. This is a major reason that also in Europe the defined benefit scheme are replaced by defined contribution ones. Given that the risk is

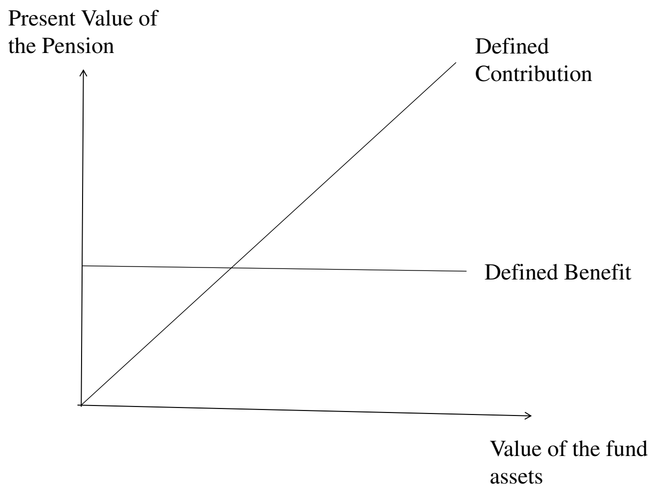


Figure 1.2: The present values of the two types of plans as a function of the pension fund asset value. The figure shows that in the defined contribution case the risk bears is the employee and in the other case the employer.

transferred to the employees the question is how stable is the market risk distribution over time. While risk of stocks is in the short run not predictable if one introduces demographic structures in the pricing of assets in equilibrium stock prices become to a significant degree predictable. A key driver for the prediction are the relative numbers of individuals in a population who are at the different stages of their life. The result intuitively follows from the stable investment and consumption pattern over the life cycle: People borrow when young, invest for retirement in middle age, and disinvest in retirement. Geanakoplos et al. (2004)[3] The authors divide the us population of the last century in five cohorts of twenty years length. They then study if predictable changes in demographic structure can lead to predictable future change in asset prices–and how significant such prices changes can be. They use an overlapping generation model, i.e. a model where the agents born at different dates overlap and where the general equilibrium of the economy is determined. The variations in their price-earnings ratios approximate the observed ones in the U.S. over the last 50 years. A substantial fall in the price-earnings ratio is likely in the next 20 years (seen from 2000-vista time) due to the baby boomer generation starting to retire. In this sense the risk transfer to defined contribution plans coincides with a period where due to the demographic change price-earnings ratios are decreasing. How could agents hedge against this risk? They could for example invest in stocks of economies which do not face the demographic risk. But workers or employees face strong firm and home biases: They hold lots of company stocks and think

[3] Other work in this direction are Abel (2002), Kuznets (1958), Mankiw and Weil (1989), Poterba (2004), Malmendier and Nagel (2011).

it's safe, see Mitchell and Utkus (2004). An other solution are risk sharing contracts between different generations in different economies, see Padovani and Vanini (2011). But such contracts do not exist, i.e. the market is not complete. Finally, one could imagine long term put options. But again the long term put options either not exist or they are not liquid. This then leads to large bid-ask spreads which makes investments too costly.

Incompleteness roughly means that there are more states than independent assets or risk sources to span these states, i.e. to generate arbitrary payoff profiles. This leads to limited risk transfer **across** states at a given date and **between** states at different dates. People cannot distribute consumption or investment at a given date or they face constraints in savings and cumulation of wealth over time. A market is **complete**, if any payoff can be **replicated** by the existing financial products spanning the financial markets. Replication means that one can form a portfolio of financial instruments such that the payoff of the portfolio equals the payoff of the replicated product at each date and in each possible future state of the world.

But financial innovation often follows from regulatory interventions into a complete market structure: The intervention either makes the market incomplete by eliminating products spanning states of the world or they transform the costs of products and their hedge. We consider regulation and the possible impact on innovation in Section 1.2.

1.1.1 Market Evolution

What is a financial market? Financial markets undergo different maturity levels reflecting different states of completeness, see Figure 1.3.

In a simple model there are three states:

- **Back-to-back**. At this earliest stage there is no liquid market. Banks can and will not enter into a risk taking position. They will search for two parties which match. One position is long and the other one is short.[4] The financial intermediary takes a non-risk position in a zero-sum game and matches the supply and demand side. Incompleteness occurs if there exist no counter party for a back-to-back matching. Risk transfer is then impossible.

- **Risk warehousing.** There are several counter parties and the bank takes itself risk positions through a trading desk. To enter into a risk position several requirements need to be satisfied. The bank needs risk capacity and risk appetite to use their

[4]One can be long or short a position. Being long a stock or a bond means the holder of the position owns the security and profits if the price of the security goes up. Short selling is the practice of selling assets that have been borrowed from a third party with the intention of buying identical assets back at a later date to a lower price to return to the lender. The short seller borrows shares and immediately sells them. He makes a profit is the stock price decreases. Shorting also refer to entering into any derivative or other contract under which the investor profits from a fall in the value of an asset. The possibility to sell short and in particular to sell 'naked short', i.e. selling the asset without borrowing the security, is in the focus of many regulatory initiatives.

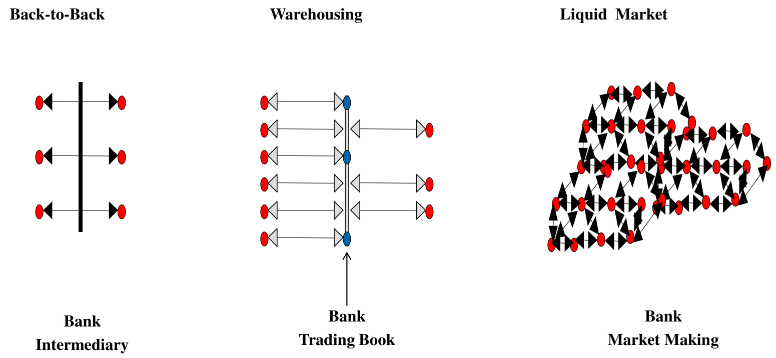


Figure 1.3: Evolutionary states of financial markets.

balance sheet for trading. If both are given, the trading desk receives a risk capital allocation and a charge for this capital. Besides the risk governance, investments in human capital (traders, sales, structurer, mid-office, back-office, risk function, operations) and IT are needed. In such a market a first generation of financial products exists but the market is not yet liquid. Financial market liquidity has two dimensions:

- Time. Products can be exchanged or traded at any time.
- Price and Volume. The impact of any meaningful[5] trade size on prices is negligible.

Financial intermediaries start their market making function, i.e. bid-ask rates/prices are set with appropriate sizes. Additional to the back-to-back status there exists a chance that a bank enters into risk sharing with a single counter party. Although the bank takes risk, this type of trading is not proprietary trading since the risk follows from clients demand to trade and is not a stand alone decision of the bank to put a fraction of their capital at risk. A banks **trading activities** can be classified in four types. First, the bank acts only on a best execution basis: A client buys a stock via its bank which executes the transaction using a stock exchange or a broker. This **agency trading** has only operational risks - the bank faces neither market nor credit risk. The next type, **principal trading**, was described above

[5]The ratio of the trade volume has to be small enough compared to the average traded sizes.

under risk warehousing. The clients demand for trading make it necessary for the bank to face market and credit risk. The third type is **proprietary trading**: The bank decides to enter into positions without any client flow behind it. Such trading acts therefore fully and direct on the capital of the bank. The Volker Rule and the Dodd Frank Act are two regulatory initiatives in the U.S. which focus to ban proprietary trading. The final type is **investment trading**. This type of trading is of a buy-and-hold type which is due to strategic reasoning, ownership of investments, joint ventures, etc. Investment trading fully affects the bank's balance sheet.

- In a **liquid** market supply and demand match automatically. The financial intermediary plays the market making function to ensure that at each date and theoretically under any circumstances liquidity is provided. The intermediary earns money by managing the **flow**. The market is complete in the sense that every payoff which can constructed in the respective market is liquid for all market participants to share their risks. We consider the creation of an option trading book in a liquid market. Consider the liquid stock Holcim. We start with a short position of 1000 calls on Holcim with price 7.232 CHF. The option price is calculated using Black and Scholes model. If Holcim stock moves, up to first order[6] a loss of CHF -587 on the derivative position follows, see Table 1.1. This sensitivity is called Delta in option pricing. The profit and loss so far is zero. To reduce risk (the Delta), we next invest in the stock. We buy 620 Holcim stocks at the price 80. The Delta of a stock is one.[7] Hence a Delta of $+33$ remains. This means that an increase of Holcim stock price by 1 CHF leads to a gain of 33 and opposite if the price falls. Profit and Loss is still zero. To generate P&L different possibilities exist. First (step three) one sells the options slightly at a higher price than their value is. This gives a profit and loss of CHF 268. Second, price movements as described above lead to profit and loss (step four where Holcim gains 1).

  Step 5 describes how volatility movements generate profit and loss. Since call options have an asymmetric payoff profile their option price is sensitive to changes in volatility of the underlying. The first order sensitivity w.r.t. volatility is called **Vega**.[8] We assume that the portfolio V is Delta neutral, i.e. $\frac{\partial V}{\partial S} \sim 0$ see Table 1.2. Volatility is 20%. If volatility increases by 1 volatility point, the bank loses 304 CHF and the opposite occurs if volatility drops. Intuitively, if volatility increases the chance that the call option is in the money increases (the value of Holcim stock is larger than the strike price). This makes a call option more expensive. Since it is a short position, an increase in volatility implies larger losses. If the trader hedges the Vega exposure he needs to trade in different options. In step 6 it is shown that if he trades in a second option, Vega of the position is reduced but Delta increases away from zero. Is it possible to control both sensitivities? Yes. Risk control means controlling first or second order sensitivities. But sensitivities are

[6] The first order option price $C(S)$ sensitivity w.r.t. to the underlying S (Holcim) is given by $\frac{\partial C}{\partial S} =: \Delta$.
[7] The Delta is the first order derivative of a position. The derivative of the stock S w.r.t. S is 1.
[8] $\frac{\partial C}{\partial \sigma} =:$ Vega.

Step 1	Product	Size	B & S	Tr.Pr.	Pos.Val.	Delta	P & L
	Option Holcim	-1000	7.232		-7232	-587	
	Position				**-7232**	**-587**	**0**

Step 2	Product	Size	B & S	Tr.Pr.	Pos.Val.	Delta	P & L
	Option Holcim	-1000	7.232		-7232	-587	
	Stock Holcim	620	80	80	49600	620	
	Position				**42368**	**33**	**0**

Step 3	Product	Size	B & S	Tr.Pr.	Pos.Val.	Delta	P & L
	Option Holcim	-1000	7.232	7.5	-7232	-587	268
	Stock Holcim	620	80	80	49600	620	0
	Position				**42368**	**33**	**268**

Step 4	Product	Size	B & S	Tr.Pr.	Pos.Val.	Delta	P & L
	Option Holcim	-1000	7.232	7.5	-7232	-587	268
	Stock Holcim	620	81	81	50220	620	0
	Delta						33
	Position				**42988**	**33**	**268**

Table 1.1: Positions in the option portfolio construction. Tr.Pr. means Trading Price, B & S the theoretical Black and Scholes model price and Pos.Val. Position Value.

mathematical derivatives, i.e. the sensitivity of a portfolio is equal to the sum of the position sensitivities. The requirement to be say Delta neutral, Gamma neutral, Vega predefined, etc. then leads to a system of linear equations. If one adds enough linearly independent products in the portfolio, the system has a unique solution.

Step 5	Product	Size	Price	Pos.Val.	Delta	Vega in CHF
	Option Holcim	-1000	7.232	-7232	-587	-304
	Stock Holcim	588	80	47040	588	
	Position			**39808**	**1**	**-304**

Step 6	Product	Size	Price	Pos.Val.	Delta	Vega in CHF
	Option Holcim	-1000	7.232	-7232	-587	-304
	Stock Holcim	588	80	47040	588	0
	Option Holcim 2	400	7.232	2893	235	122
	Position			**42701**	**236**	**-182**

Table 1.2: Position in the option portfolio construction. The figure Delta is expressed in numbers of Holcim shares.

Besides market evolution from a **risk and activity** point of view, the evolution can be considered from a **technological** side. From the first open outcry markets, the technological revolution of market places started in the 70s/80s of last century. In electronic markets it is possible to trade longer, trade smaller volumes and trade faster. Automatic trading has different operational risks than traditional trading.[9] The probability that events realize are smaller in the automated world; but is hard to compare the severity of operational risk events in the two trading approaches. We consider an example.

The first trading day of the **Facebook IPO** stock listing in 2012 at NASDAQ showed that realization of operational risk in electronic trading can have a serious severity (i.e. loss and reputation) impact. Although NASDAQ simulated higher trading volumes than actually occurred they were unprepared for increasing numbers of cancelled orders in the hours leading up to Facebook's debut. This caused a 30-minute delay in the start of Facebook trading.

Automatic trading raises some questions. First, one often speaks about **algorithmic trading**, electronic trading and **high frequency trading**. How are they defined, who are the players, how do they interfere with the functioning of financial markets? Second, how does regulation boosts or constrains these types of trading? How does the reduction in risk capacity of the trading units in investment banks changes their business model? We consider this in more detail in Section 1.3.

Given that markets are incomplete, equilibria are in general not Pareto optimal. One might think that there will be a demand for increased opportunities to share risk

[9] Operational risks are risks due to failure, non-availability, error, fraud of people, systems and processes.

which then increases welfare of everyone. While this is true for **single** consumption good economies it is not true for more complicated ones when **not all** missing markets which complete the original incomplete market are added - by the First Welfare Theorem adding all missing markets cannot lead to an overall reduction in welfare. Elul (1995) shows that the addition of a new security can have almost arbitrary effects on agents' utilities. The introduction of a new security can generically[10] make all agents strictly worse off or strictly better off or favor any group of agents over another. With several goods there exist complicated feedback effect which counteract the intuition of the single good economy. The result, however, does not states how difficult or easy it is to introduce an asset which reduces or increases welfare of all agents in the economy.

We discuss innovation along several dimensions. We follow Tuffano's (2002) structure.[11] A must-read on Innovation is the book of Allen and Yaglo (2010).

- Information Asymmetries.
- Costs (transaction, search, marketing).
- Taxes and regulation.
- Globalization and innovation.
- Technological shocks.
- Who innovates?
- Impact of financial innovation on society.

We only sketch the asymmetric information and cost issues and refer to the paper of Tuffano (2002) for details. The role of **information asymmetries** and the theory of contracting or security design which analyze them changed during the last decades. The primary role of innovations forcing the revelation of information switched to innovations exploiting low information costs. A second observation is that contract theory provides a rationale for some basic contracts such as equity or debt but not for more complicated structures. Is there no rationale for such structures or are the methods of contract theory to narrow to allow for them?
Costs and in particular **transaction costs** could be lowered by factors up to 100 in the past due to innovations. As with information asymmetries further reducing transaction costs is less important for liquid assets. Other characteristics of products and services such as flexibility, customized design, low balance sheet impacts, compliance and transparency are becoming more important.

[10] The technical term 'generic' means that a property is stable and holds in almost all cases.

[11] Geanuracos and Millar (1991), Walmsley (1988) and Crawford and Sen (1996) are popular business books about innovation.

1.2 Taxes and Regulation

Miller (1986) states: *'The major impulses to successful innovations over the past twenty years have come, I am saddened to have to say, from regulation and taxes.'* The list of tax and regulatory induced products includes zero coupon bonds, Eurobonds, equity-linked structures and trust preferred structures. The search to maximize after-tax returns has stimulated much innovation. An analysis shows that in the U.S. taxes where yet a driver for innovation a century ago Kane (1986) identified what he calls the "regulatory dialectic" as a major source of innovation. Innovation responds to regulatory constraints, which in turn are adjusted in reaction to these innovations. Bank capital requirements impose costs on the affected banks. They then use innovation to optimize capital charges in light of these constraints. The Eurobond market starting in 1966 was motivated by regulatory concerns to circumvent reserve requirements. Given the many regulatory initiatives after the financial crisis and the initiated innovations we can simply restate Miller's statement to hold true at present.

We can understand the impact of regulation on banks more formally. Let ϕ be strategy a bank chooses to optimize its value function $V(\phi)$. The optimization is over all strategies which are admissible, i.e. $\phi \in \mathcal{A}$ with $\mathcal{A}$ the admissible set. This set is the intersection of exogenous and endogenous constraints such as budget constraints, liquidity constraints, risk constraints, market access constraints. Any new **regulation** leads to a new set $\mathcal{A}' \subset \mathcal{A}$. F**inanical liberalization** means $\mathcal{A} \subset \mathcal{A}'$. The optimization of the value function given the smaller constraint set automatically implies a **lower or equal** optimal value. Every new constraint has a positive shadow price. This holds for a single bank in the regulation case and correspondingly in the liberalization case. If we consider a system of banks new constraints do not necessarily have a uniform decreasing impact on all bank's value function. Typically, new regulations induce shifts in market shares for the different banks: Although the restriction leads to smaller constraint sets for all banks some end up with a larger optimal value since their market power increases compared to other ones. The central counter party (CCP) issue is an example. This initiative makes over-the-counter derivative trading less profitable if one does not clears the trades in a central counter party (clearing house). To do this, one banks need access to the central counter parties. This requires technology and a size of the trading unit which medium or smaller sized banks do not meet - they have to clear their OTC contracts via a larger bank in a clearing house. This provides earning to the larger institutions which offer the clearing service to smaller ones. For these institution the net effect of additional earning minus the cost of the new constraint might be positive.

1.2.1 Taxation

This section is based on Glauser (2011, 2012). Taxation is linked to tax law, accounting law, finance and corporate law. The levying of tax requires a **legal basis**: The law must clarify the tax subject, the tax object, the taxable basis and the tax rate. We consider taxes in the sequel which are contributions paid without any specific consideration. Direct taxes mean that the taxpayer and the tax bearer are the same (income tax, wealth tax). If the two types disagree, one speaks about indirect taxes (VAT, stamp tax). Income is typically taxed in two ways. Ordinary income taxation means that the taxes are due to the person receiving the income. On the other side, withholding taxes (WHT) means that taxes are due to the debtor paying a specific item. WHT is frequent on salaries, dividends, interests and royalties. The legal form of companies, i.e. subsidiaries vs. branches, has important impact on the WHT. While branch offices and the head office form a single legal entity and no WHT is due for dividends etc. which flow to the head office, in case of subsidiaries all entities are legal entities which pay the WHT. The WHT serves as a guarantee for regular declaration of income by residents and it levies a tax on income by non-residents. Figure 1.4 illustrates different taxations for Switzerland.

	Income Tax	**Profit Tax**	**Net Wealth Tax**	**WHT**	**Stamp Tax: Issuance**	**Stamp Tax: Turnover**
Tax Subject	Individuals liable pay tax in CH	Corporations liable pay tax in CH	Indiv./Corpor. on net wealth in CH	Payer of taxable income. Transfer to the beneficiary	Shareholder contribution into corporation	Transfer of ownership
Tax Object	Enrichment during tax period	Enrichment during tax period	Indiv: Wealth Corp: Equity	Gross amount of invest. Income	Entity receiving the contribution or issuing the bond	Securities dealer involved in the deal
Taxable Basis	Private: All type of income but not capital gains	Taxable profit determined by accounting law	See Tax object	See Tax object	Fair market value or for bonds % p.a. of bond duration	Consideration
Tax Rate	Progressive in % of taxable basis	Fedearl: 8.5%. Varies between Cantons	Varies between Cantons	35%	1% for contributions Bonds more complicated	0.15% of consideration for Swiss securities 0.3% for foreign securities

Figure 1.4: Taxation examples for Switzerland.

The main applicable tax law is the domestic one. Since most international tax law relies on the concept of residency and not nationality (an exception is FATCA, see below), domestic laws applies. Most important international tax laws are double tax treaties (DTT). The OECD contract are used for most DTT. DTT cover direct taxes but not VAT. Since they are an allocation tool, there is no legal basis: Domestic law applies for the effective taxation.

We consider tax planing strategies for **private investors**. The first distinction in taxation is whether the funds are declared or not declared. If they are not declared, no taxes are due and no tax planning is necessary. But today taxation and penal consequences for the clients, the banking institution and for the bank's wealth managers are much more severe than in the past. What are the reasons for these changes? First, the statements are true for western countries. Second, although penal consequences changed drastically, it is doubtful that this is due to a change of social or political norms. Often a government does not accept tax evasion of its citizen but may well be indifferent if their jurisdiction allows citizens of other countries to evade taxation. It is not always the norms which changed but the perception of norms of other nations and the determination to protect a nation's norm changed. The U.S. for example started in the 70s of last century to make sure that all U.S. persons pay taxes, independent where they were living. In 2001, the IRS[12] introduced the qualified intermediary program, see Section FATCA for details. This program showed some loopholes which IRS wants to close using the program FATCA. That is, to define taxation justice for U.S. persons worldwide is not new to the U.S. policy although for many non U.S. financial intermediaries, which are affected by FATCA, this might seem a new initiative. Different is the situation for the Eurozone. There tax evasion is a much less severe punished activity which underwent a regime shift after the financial crisis in 2008: Tax evasion in many countries of the Euro zone is now pursued with a higher determination than before the crisis. Given the economic status of the Euro Zone there is no need for policy makers to increse the awareness in the population - it simply pays for the policy makers if they prove to restore taxation fairness by pursuing tax evaders. Third, although in the past not declaration of money was tolerated to a larger extend this is no reason per se why this should continue in the future. Which of the explanations, including others, is the most important one to explain the changes is irrelevant for us since we focus on **declared** funds only, i.e. where tax planning is a crucial discipline in wealth management.

Tax planing has three main streams, see Glauser (2012) for further details:

- Choose the right tax jurisdiction.
- Choose the right products.
- Choose the right vehicle.

1. Choose the **right tax jurisdiction**. A first distinction is between unlimited and limited tax liability, i.e. the obligation to pay taxes on worldwide income/net wealth or whether taxation is limited to a specific presence in a jurisdiction. The first one is linked to the physical presence in a country or to the nationality (U.S.). Figure 1.5 illustrates some cases. Exceptions are domestic law which for example applies to foreign real estate exempted or double tax treaties. In any case, the location of bank account is irrelevant, it is the residency of the tax payer which matters mostly. One next has to distinguish between income and withholding taxes. The first one are due by the person receiving the income, the second one frequently applies on dividends, interests or royalties. Concerning the withholding tax one has to make sure whether the tax is due by company distributing dividends or by borrower

[12] The IRS is the U.S. government agency responsible for tax collection and tax law enforcement.

paying coupon on bond. or whether the tax is due by the paying agent. In the first case, the location of the bank account is irrelevant. Tax planning is achieved by transferring the taxpayer, in the second case the location of the bank account is relevant and tax planning is achieved by transferring the bank account. Since tax payers can improve their tax situation by moving to another tax jurisdiction the question is, what determines good alternative domiciles? Besides preferences where to live other taxation issues are: Age of the individuals, type of taxation system in the potential new domiciles, what are the sources of income, are the double taxation treaties to avoid double taxation, what is the impact on the asset allocation, i.e. shifts between real estate asset class and traditional assets class, capital gains vs. other incomes?

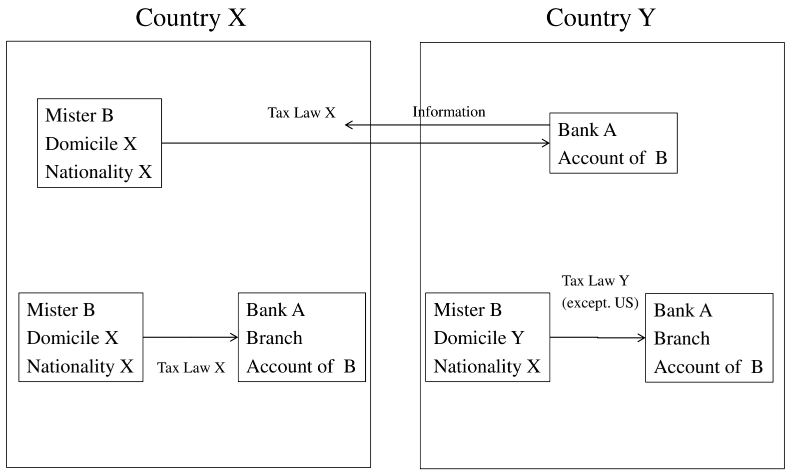


Figure 1.5: Relationship between domicile, nationality, bank account location. A major discussion between the countries is the first case. If in this case Bank A in country Y does not deliver information to the tax authorities for country X and if the client does not want to declare his wealth or income tax losses for country X follow. The debate is to force banks in country A to provide information to country X's tax authorities.

We consider double tax treatises (DTT) in more details. We state or recall, that (for) Double tax treaties (DDT):

- overrule domestic law.
- function as an allocation tool, i.e. which state can tax what type of income?
- are useful to avoid double taxation.
- require that a person is resident at least in one of the two countries.

DTT have further objectives:

- Legal and administrative cooperation, i.e. country X asks country Y for information about a citizen of country X. Such requests typically break banking secrecy (here in country Y). The DTT defines the information which is exchanged.
- DTT define the tax jurisdiction which applies to specific individuals.
- DTT clarify what happens when two states can tax a part of an individual's wealth or income?
- DTT clarify what happens when there are conflicting views between the two states?

As an example consider an individual investor, with domicile in Switzerland, which owes stock of an U.S. based company. The company pays a dividend of USD 1'000 in 2010. The following table shows the tax calculation under a DTT agreement between the USA and Switzerland.

Taxation U.S. and DDT	
Gross Dividend	1000
Withholding Tax DTT US-CH 15 %	-150
Net Dividend	850
Income Tax CH 25 %	250
Remaining Taxable Amount CH	100
Total Tax on Dividends	
Withholding Tax DTT US-CH 15 %	150
Income Tax CH 25 %	100
Total Taxation	250

2. Choose the **right products**. A main issue are whether payments arise from investment vehicles (dividends, coupons) or from a third party such as buyers of shares or bonds. In the first case the withholding tax applies; in the second case there is in many jurisdictions no withholding tax on the capital gains. Since in many jurisdictions capital gains are taxed with lower rates one tries to convert investment income into capital gains. One simply can sell bonds or share before they pay the coupons or dividends, one can reinvest the dividends or structure the products such that mostly capital gains realize. But such type of transactions are limited by law and local countermeasures exist. A particular strategy regarding is **dividend washing.** Consider an foreign investor in a Swiss stock paying dividends. The foreign investor faces the withholding tax which is not refund but for a Swiss investor, if he declares the investments, the WHT is refund contrary. Therefore the foreign investor could sell the stocks just before dividend date to a Swiss investor. This then declares the dividends and gets the WHT back. He then sells the shares back to the foreign investor shortly after dividend date. This kind of transaction is

illegal in Switzerland: Tax authorities will not refund the tax amount to the Swiss investor since he fails to be the **beneficial owner**. Another form of conversion was to use insurance wrappers: Client invests in a life insurance. Funds are however invested by the insurance company in a specific bank account. Client can influence the investment strategy. If payment is qualified as insurance payment: different tax rates or even tax exemption follows. Besides conversion, deferral is a product strategy since the moment of taxation can differ between products and tax jurisdictions. In Switzerland: most investment funds are treated as transparent entities, i.e. income is taxed when generated and not when distributed and capital gains are not taxed.

3. Choose the **right vehicle**. The main question in Switzerland is: Do the investor keep control (directly or indirectly) over the funds? In case of transfer to a third party (trust or foundation), the transfer of ownership is possible but (in Switzerland) is looked at from an effective / economical point of view: The settlor is not a beneficiary of the trust / foundation, the rust (foundation) is irrevocable and discretionary; and tax authorities look at all documents (letter of wishes, etc.). If the transfer is effective, the transferor does not need to report assets / income in his tax return and no inheritance tax is due. Contrary, if the transfer is not effective he needs to report assets / income in his tax return and inheritance tax upon death becomes effective at this point of time.

1.2.1.1 FATCA

The Foreign Account Tax Compliance Act (FATCA), a channel to finance the Hiring Incentives to Restore Employment Act, is a tax initiative of the U.S. which affects ten thousands of financial intermediaries in the world. What is the rationale? It is estimated that out of 7 million U.S. Persons living outside the U.S. more than 6.5 million have never filed a U.S. tax return and U.S. residents have undeclared offshore accounts (Ernst&Young (2012)). Why? First, there are U.S. persons using legal vehicles to hide that they are U.S. persons. Since there is no look through by the IRS, these individuals will not pay taxes. Second, there are U.S. individuals which only invest in non-U.S. assets. Since the 2001 initiative by the IRS only focus on U.S. persons holding U.S. assets, the IRS intends to close the two mentioned loopholes, see Figure 1.6. More precisely, after 2001 the IRS has relied on non-U.S. banks and other **Qualified Intermediaries (QI)**[13] to enforce compliance with specific U.S. withholding tax and information reporting requirements. Under a QI agreement with the IRS the QI assumes certain documentation and withholding responsibilities w.r.t. U.S. account holders and U.S. source income. The IRS in exchange permits the QI to certify the status of its non-U.S. account holders using

[13] The term refers to a foreign bank that has entered into a special agreement with the USA IRS to report information to the IRS on Americans with accounts at the bank. The bank also has a requirement to withhold a percentage of the interest income paid to the account. The bank has an obligation to file tax reportage forms to the IRS. If the bank fails to comply they can be prosecuted criminally. In 2005 there were about 5,000 banks participating in the program.

anonymous pooled reporting without revealing the identity of the non-U.S. clients.

	U.S. Person	Non-U.S. Person	"Non-U.S." entity with U.S. ownership
U.S. Securities	W-9 1099-reporting	Double tax treaty access through pooling	In scope of FATCA
"Non-U.S." Securities	In scope of FATCA	Currently out of scope	In scope of FATCA

Figure 1.6: The IRS has information about the upper left cell. With FATCA, the IRS asks for information about three more indicated cells. *Source: Ernst&Young (2012).*

Given that U.S. persons used legal constructs to hide that their are U.S. persons, IRS would like to know **all** investments (US source and non U.S. source) and **all** revenues and proceeds from investments undertaken by a U.S. person. They need to be reported to the IRS, **irrespective** of the type of holding and booking location. That for foreign financial intermediaries have to find **any** American account holders and disclose their balances, receipts, and withdrawals to the IRS or be subject to a 30 percent withholding tax on income from U.S. financial assets held by the banks. Owners of these foreign-held assets must report them if they are worth more than USD 50'000. Account holders would be subject to a 40 percent penalty on understatements of income in an undisclosed foreign financial asset. Figure 1.6 shows that U.S. accounts can belong to natural persons as well to legal entities which are US-owned, i.e. more than 10 percent of the entity is owned by U.S. persons.

For the foreign financial institutions the fulfillment of the FATCA requirements has four main steps: First, identification of U.S. accounts. Second, a waiver by U.S. client. If the U.S. client is recalcitrant the **financial intermediary (FI)** will terminate the client relationship. Else, step three, annual reporting to the IRS takes place. The fourth step are penalties for non-participating foreign financial institutions and recalcitrant account holders. Figure 1.7 shows the different parties and information flows. At the beginning

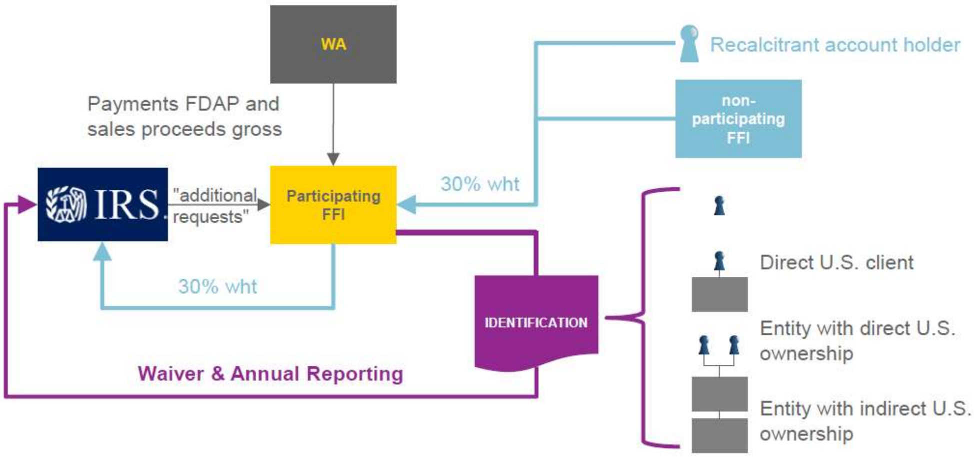


Figure 1.7: FATCA on one page. FFI are the Foreign Financial Intermediaries, i.e. one the estimated 100'000 intermediaries which need to apply FATCA, WHT means Withholding Tax, FDAP means fixed or determinable annual or periodical income and WA is the withholding agent.*Source: Ernst&Young (2012).*

are payments FDAP[14] which a withholding agent makes to a participating ('good') **foreign financial intermediary (FFI)**. The FFI identifies the clients. For those which are recalcitrant ('bad ones') the FFI transfers 30 percent of the income to the IRS. The other ones are reported to the IRS.

One complexity for the FFI is the search for U.S. persons, i.e. what defines an U.S. person? They have to search for U.S. residents/citizen, i.e. people with an U.S. passport. But also individuals with an U.S. place of birth, U.S. address, U.S. mail address, U.S. phone number, standing order in the US, c/o or holdmail as sole address need to be

[14]FDAP means fixed or determinable annual or periodical income.

identified. If a person is indeed a U.S. person, the IRS requirs to fill out the W-9 document[15] and a waiver. If the person is **not** an U.S. person, the person has nevertheless to document their non-U.S. status. Some persons will be recalcitrant to such requirements. But this can be costly.

Consider the case of a recalcitrant bank client which is a non-US person. The recalcitrant client has an account of CHF 100'000 within the bank which itself participates as a FFI. Then the passtrhu concept applies, i.e. one assumes that a percentage of the 100'000 CHF are invested in U.S. and non-US assets. Assume that 20 percent is invested in U.S. assets: the **passthru payment percentage** (PPP). Then FATCA charges 30 percent on the interest income of 20 percent of the assets. If total income is 2'000 on the 100'000 the WHT is CHF 120.[16] But for the recalcitrant client the story does not end here because the PPP can increase and hence the WHT. That for assume the following relationships of the participating FFI to other three other FFI where the following PPP apply to the 80'000 CHF (the 20'000 are yet attributed to he participating FFI):

- One to a bank (FI) which is participating with a PPP of 50 percent. Of the 80'000 non-US assets, 40'000 are invested in this bank.
- One bank fails to provide the PPP, i.e. since the PPP is not known, the PPP of 100 applies. 30'000 are invested in this bank.
- One bank is non-participating FFI, i.e. the PPP is zero. 10'000 are invested in this bank.

For the original bank a PPP of

$$\frac{20'000 + 1/2 * 40'000 + 1 * 30'000 + 0 * 10'000}{100'000} = 70\%$$

applies. Therefore, the recalcitrant client has to pay the 30 percent withholding tax on CHF $1'400 = 0.7 \times 2'000$ income, i.e. he pays CHF 420.

1.2.2 Investor Protection

Financial institutions face **conduct-of-business risk**. The FSA for example charged fines of about 50 Mio. Pounds to different institutions in 2011-2012 without a 290 Mio. Pound fine for attempting to manipulate LIBOR. The conduct-related fines can be classified into the categories 'product governance' (Failure to inform clients), 'fair customer treating' (Excessive charges) and 'client assets' (Failure to protect and/or segregate client money). What is an investor? What are investment firms? From an economic and legal perspective investment firms can

[15] The form W-9 is a tax form of the IRS for taxpayer identification certification. It is used by third parties to collect identifying information to help file information returns with the IRS and it helps the identified person to avoid backup withholding.

[16] 20 percent times 30 percent of 2'000.

- execute orders from investors,
- provide advisory to investors,
- manage wealth of investors in a delegated form.

Central for investment firms operating in the Euro Zone is **MIFID**[17] (Market in Financial Instruments Directive). This directive allows banks, brokers, and other financial intermediaries to provide financial advisory in the European Economic Area (EEA) without the need to comply with the local regulatory requirements (consistent market access), they can sell in the EEA products and services without the local authorization and the directive sets forth prescriptive obligations upon the firms considering organization and conduct of business (investor protection rules, improved transparency in trading, increased product reporting and documentation, limitations of inducements such as kick-backs). MIFID categorizes the investors as follows: Investment firms must define written policies and procedures to the categorize their clients:

- **Eligible counter parties**, i.e. institutional clients such as investment firms, pension funds, national governments.
- **Professional clients**. A client classifies as a professional one if he possesses experience, knowledge and expertise to make its own investment decisions and properly assess the risks that he incurs. A client can be classified a professional one per se or at their request. Per se applies to entities which are required to be authorized or regulated to operate in the financial markets and which satisfy a certain size (for example a balance sheet which is not less than 20 Mio. Euro). A client classifies as professional on request if he shows an experience and activity level in the financial markets in the near past, if the value of cash deposits and financial instruments exceeds a half-million Euro and if he worked in the financial sector. If two out of the three criteria are met classification as a professional investor is possible. We note that the standard segmentation approach using wealth as unique variable is not sufficient.
- **Retail clients**. By MIFID, all other clients are retail clients.

The investment firms have to notify the clients of their right to request a different categorization and clients can both, opt up and down, i.e. choose a less or more severe protection category than the bank itself would define.

Given how in the EEA client protection is regulated by the MIFID, several questions arise. First, why should one protect investors? Second, how can regulation with the goal of investor protection be defined? Third, how can a theoretical client protection concept be implemented to act effectively?

[17] There are two directives, MIFID I and MIFID II. MIFID II will be effective July 2015.

We consider the first question. If investors are not protected but think they should have been they will stop to invest. But if there are no investors there are no financial markets. Hence it is also in the self-interest of finanical intermediaries to possess efficient and effective client protection procedures. Financial instruments are different to many other goods. We compare the decision problem to buy a car and to invest into a simple derivative, say a call option on Nestle. Cars are often termed experience of goods whereas the call option is a credence good, i.e. a good the agent has to believe in. Clearly, he also has to believe in the car but as a physical good the agent can value without risk today many characteristics of the good - functionality, design, security standards and so on. The time when the agent can value the option lies in the future and is uncertain. To buy the option today requires the client to understand what will happen to the option under different scenarios. This is a much more abstract exercise and also bears much more risk for the seller (bank) than for the car seller: The client relationship manager has to show the financial consequences in a as comprehensive way to the client. The new technologies allow to address this question radically different than in the past, see Section 1.3.1. Clients also need protection since they often do not have the means to validate the products. This can be missing data (where should a retail client obtain a correlation parameter?), knowledge or time. Comparing the call on Nestle with the car case, in the latter one a driving licences is necessary and sufficient to buy a car from a legal perspective (unless other legal restrictions apply to a specific person). For the financial instrument so far a test where the client proves that he understands the product is not required, at least not for retail clients. Contrary to a driving licence such a test turns out to be more demanding and needs to be updated on a regular basis given the financial product innovations. Financial intemediaries often reject the idea to ask for tests for their clients since they fear that most clients will not be able or willing to pass the tests. As a consequence a large part of the product offering can no longer be offered to a assumed large fraction of retail clients.

A major concern why clients need to be protected are **conflicts of interest**. It is difficult without any protection for the client to control that the products offered are the best suited for him and not for the benefit of the relationship manager. The client does not knows the incentives of the relationship manager and he does not know whether there is any strategy behind the product offering. But this also applies to other types of products: We do not know the incentives of a car dealer in the above example neither. Contrary to the financial situation we accept that this might be conflicting. Although it is evident that it is impossible to have a long term client relationship if the bank acts not in the client's interest the history of banking or investment is full of examples where ruthless advisors or such with a short horizon did not acted in their client's interest. The difficulty with investor protection is that the advisor has to make transparent the product, has to make transparent his incentives, has to elicit the knowledge and experience level of the client and finally process all this information such that the output is that a product is suitable for a client or not. This kind of view which is in some form or other underlying many initiative for client protection is likely to fail since it is too demanding to be implemented. We discuss in Section 1.3.1 how using new technologies some of the above duties or responsibilities can be changed: The client can explore the products himself and then decide whether a product is suitable or not for him. The client relationship manager takes in this approach the position of a coach.

How can regulation be defined? There are self-regulation and regulation by the state or regulators, direct and indirect regulation, principle based and detailed regulation. While there are many pros for self-regulation, the recent banking history shows that the interference of compensation incentives and self-regulation destroyed in many jurisdiction the legitimacy of this approach. In other words self-regulation is more and more replaced by ordinary regulation. A part of regulation is **suitability**, i.e. the test of a financial firm, when advertising a retail client to purchase a particular financial instrument, that the product is appropriate for the client. That is the product or service offered should match the client's financial situation, investment objectives, level of risk tolerance, financial need , knowledge and experience by the Basel Committee on Banking Supervision (2008). To provide an appropriate suitability process **conduct of business rules** are required: Firms must ensure that they act honestly, fairly and professionally in accordance with the best interests of the investor and treat investors fairly. Furthermore, the rules regulate the behavior of the service provider and they ensure that the clients do not suffer from the position of strength of the financial intermediaries and from asymmetry of information. Both, suitability and conduct of business rules are key parts of investor protection.

Following Oliver Wyman (2012) the following requirements are key in successfully managing conduct of business rules:

- Specification of stakeholders roles. A common model of risk management, the three line of defense model, is applied here too. The 1st line where the client relationship manager acts needs to act such that the client is treated accordingly to the regulatory requirements. This is implemented using appropriate incentive schemes, compensation schemes and selecting people which act in line with the bank's culture. The 2nd line defines the guidelines, monitors risk and tracks corrections. This is the risk and compliance line of defense. Employees of this line need to be separated from the first line w.r.t. their reporting lines and the incentive as well the compensation schemes have to be disjoint. The 3rd line provides controlling assurance. They review the activities of the first line.

- Firm processes for product approval and marketing material. This requires the definition of incentives, compensation and a culture of product design which is appropriate with the clients needs. Furthermore, the marketing material has to be balanced, complete and understandable. This requires to avoid known pitfalls from behavioral economics or finance research: Failure of people to apply the laws of probability correctly and behavioral phenomena such as anchoring, prominent numbers, overconfidence, etc. are well known, see Rabin (1998).

- Tools demonstrating fair treatment. Fair treatment needs to be demonstrated pre- and after-sale during the life cycle. Without powerful tools this is not feasible. One should allow clients to generate at any time a full fledged report about their portfolio. This report should state what the products are, how did they behaved in the past, a profit and loss decomposition, etc. This provides transparency but is this sufficient for fair treatment?

- High data quality and robust process to satisfy client assets requirement. This requires the compilation of comprehensive and reconciled client lists, reduction of workflows and processes, control process end-to-end and readily available management information systems.

- Strong oversight over market abuse and financial crime. To insure this a basic requirement is that senior management takes responsibility, to have continuous risk assessments in place and to create clear documentation to tackle financial crime.

While most would agree with the intentions of suitability and conduct of business rules the main question remains: How and to what extend can these theoretical advices/rules be **implemented** in real life? Consider a client which today has a portfolio of securities consisting of a single stock BMW in a securities account of a bank X. What is the client's financial situation? First, many clients have more than one bank relationship, say the client has another securities account in a bank Y. Typically, bank X does not knows about this account. Second, the client is a member of a pension scheme. The portfolio composition of the client's pension assets are also not known. Third, other assets such as real estate buildings or possible legacys in the near future should be considered to obtain a view on a client's financial situation. But even if the client has only a single deposit in bank X the situation becomes quite complex if he is actually also invested in funds and structured products. To obtain a reasonable risk assessment the bank needs to unbundle the funds and structured products: They can also contains BMW stocks. This is feasible if the in the bank X securities deposit products issued by bank X are contained. But if products of other issuers are part of the deposit it becomes likely impossible for bank X to unbundle the products in their basic securities. Formally, the mentioned transparency problems can be stated as follows. Let V be portfolio of the client where

$$V = V^X + V^h$$

is the portfolio part in bank X and V^h the hidden part: Assets in bank Y or in the pension fund system of the client. Hence, V is not observable to the bank-X client advisor. Next, suppose that V^X consists of BMW stocks, funds with investment in BMW and structured products with BMW as (one) underlying value. Then,

$$V = \omega S + F(S) + C(S) + V^h$$

represents the portfolio value with S the BMW stock price, ω the observable fraction of the client in Bank X, $F(S)$ the fund value as a function of S and $C(S)$ the value of the structured products. Suppose the bank is able to determine the fraction of fund investment in BMW. This leads to ω_1, i.e. the fraction of the client portfolio value at bank X invested in the fund. Similarly, the structured products can be unpacked leading to an exposure in BMW stocks and possibly other underlying values. The total exposure in BMW can be written as $\omega_2(t)$ which consists of a direct investment in S and a replication of options on BMW in a time varying part proportional to BMW stock (Delta) plus a remaining part[18]. Summarizing,

$$V = (\omega + \omega_1 + \omega_2)S + V^X(\text{not in } S) + V^h$$

is the decomposition of the client position if full information is available about the position in bank X. To analyze this decomposition several demanding task follow. First, the fraction of wealth ω_2 is not constant over time in that part which replicates options. Second, not only the decomposition in the positions matter for future analysis but also the dependencies, i.e. how does S statistically or analytically correlates with the non-S positions.

If we consider the issue to capture the client's need similar complexity matters arise as before. To consider the needs basically means that one needs to start with an asset liability analysis (**ALM**). That is, the client needs to define the liabilities over time (education for the children, investment in real estate, etc.) distributed over time, the assets which can be used to finance the liabilities together with the expected income stream over time. Fortunately, using modern IT technologies such programs can be designed in a user friendly way. Given this information one has to analyze to what extend the ALM plan is feasible - is it feasible risk free, how much market risk is required to finance the liabilities with an acceptable shortfall, which projects does one needs to remove or postpone to reduce the financing risk. The drawback of such an analysis with a performing IT tool is model risk. The scenarios which are needed to simulate to what extend the liabilities can be financed require return, correlation and volatility estimates for the asset dynamics. This is model risk due to the bank. On the client's side the estimated future income stream can be ex post largely different than the realized one.

Conjecture 1.2.1. *The regulatory initiatives at the point of sale increase conduct of business risk to a level where it becomes optimal for the intermediary to offer to eligible and professional clients execution platforms and to use mandate contract for retail clients. The third function of advise will become under heavy pressure. The first service means that the autonomy and responsability are fully on the investors side. This service is made possilbe by developing trading and analyse tools such that the investors can act as if they were working in the bank's asset management.*

1.2.3 Regulation, Overview

Many regulatory initiatives are under way in 2012. Many of them were either triggered by the financial crisis 2008 or their actions were enforced due to lessons learned in the

[18] An option is replicated by a linear combination of cash instruments and cash.

crisis. Some initiatives in the U.S. are **Dodd Frank Act, Hire Act, Volcker-Rule**. , in the Eurozone **MIFID, EMIR** and 'world wide' **Basel III**. **Fatca** is a U.S. initiative which impacts financial intermediaries outside the U.S. These initiatives

- impact the OTC derivative markets;
- consider cross border banking;
- impose trading restrictions, i.e. proprietary trading;
- strengthen client protection and client reporting;
- strengthen the capital basis and the liquidity status of the banks.

These changes will change the landscape of banking business: Some business will be stopped, some business will be done different, some business will be done by other institutions. It will undoubtedly also promote financial innovation - a prominent one is the change in OTC business. Figure 1.8 gives an overview of some aspects of the different initiatives.

Figure 1.10 shows the regulatory changes from two different perspectives. The first one provides the **hierarchy of the policy groups**: The G20 is the leading organization concerning the regulation of the financial industry. The Financial Stability Board (FSB) is the know how center which elaborates solutions for the G20. The Basel Committee regulates in its new regulation accord Basel III the whole balance sheets of the banks. This requires not only the regulation of capital but also the consideration of the liquidity dimension. On the lowest level are the binding national regulatory bodys which consider the consultative packages and finalize them into binding law. The proposal Basel III for example has a Swiss Finish and a UK finalization. The right panel in Figure 1.10 shows the interplay between regulation and the banking industry response. It follows that most institutions act such as to **comply** with the changing regulatory environment. Only few according to Oliver Wyman head for the second step: Where are the **strategic** business alternatives due to the regulatory change? There are many alternatives. One observes that bankers are stronger to see the need and to define goals if the alternatives are in the client relationship interface than if the back or middle office functions are considered, i.e. so-called **transaction banking**. One observes more attention related to MIFID than say CCP clearing.

Although a timeline exists for the different initiatives, most of them are finalized only on a high level, see Figure 1.9 The next step of detailed rule making will lead to changes in the content of regulation and/or also to changes in the timeline. The complexity banks need to master will lead to institutions which are able to fulfill the requirements maintaining the actual business activities and organization and it will also lead banks to conclude that they cannot master the regulatory change without substantially changing the scope and organization of the institution. For small banks it might become too expensive to comply with the regulatory rules. To make all their middle and back office functions compliant is either not feasible or not economic meaningful. Such institutions

Category	Regulation	Country	Short Description
Capital & Liquidity	Basel III	27 Countries	More & better capital quality Equal weight capital an liquidity management Conservative valuation of market and counter party risk
Client Protection	MIFID / FINMA	EU-Zone / CH	Improving client protection, best execution of all client orders
Client Reporting	FATCA	World	Increasing data and reporting requirements for US business applicable to non-US financial institution
OTC Derivatives	Clearing obligation Dodd-Frank/EMIR	US / EU-Zone	Clearing of OTC contracts via central counter parties (CCP)
	Reporting obligation Dodd-Frank/EMIR	US / EU-Zone	Reporting about OTC transactions out of centralized data bases
	Trading Obligation Dodd-Frank/EMIR	US / EU-Zone	Transferring execution of standardized OTC derivatives away from bilateral form onto organized trading platforms
Crossborder Trading	Giovannini, T2S	EU-Zone	Stronger integration and standardization of the European settlement and clearing infrastructure
Trading Restrictions	Volcker Rule	US	No proprietary trading
	Position limitation and short selling restrictions	EU Zone / US	Position limitation for commodity contracts Forbidden naked short selling and uncovered CDS

Figure 1.8: Overview over regulatory initiatives. The Dodd–Frank Wall Street Reform and Consumer Protection Act is a federal statute in the United States that was signed into law 2010. MIFID (Markets in Financial Instruments Directive) is a European Union law that provides harmonized regulation for investment services across the 30 member states. EMIR (European Market Infrastructure Regulation) a major body of securities market regulation for the European Union. FINMA is the Swiss Regulatory Authority. The Volcker Rule is a specific section of the Dodd-Frank Act. FATCA (The Foreign Account Tax Compliance Act) is subtitle of the Hiring Incentives to Restore Employment Act (HIRE) in the US. T2S is a Eurosystem service for securities settlement setup by the ECB. Giovannini Group on is a group on cross border clearing and settlement arrangements in the European Union.

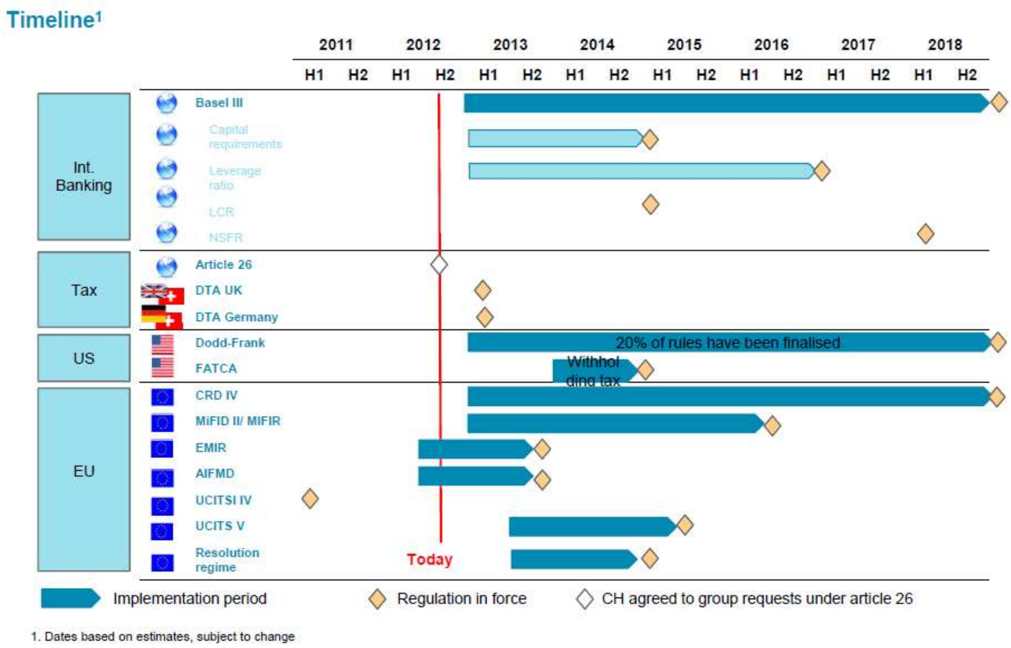


Figure 1.9: Timeline of international regulation. *Source: Oliver Wyman (2012).*

then can become part of a larger entity or outsource some functions. If we consider all activities related to the middle office and the back office as **transaction banking**, a possibility is to outsource this part of the bank while keeping the activities in the client-client relationship manager interface. Transaction management includes operations, custody, trading functions and product management among others. For Switzerland estimates are that about one-third of all banks will come to such types of conclusions. This defines a possible case for larger institutions which can offer transaction services to this smaller banks. Some questions given these initiatives are:

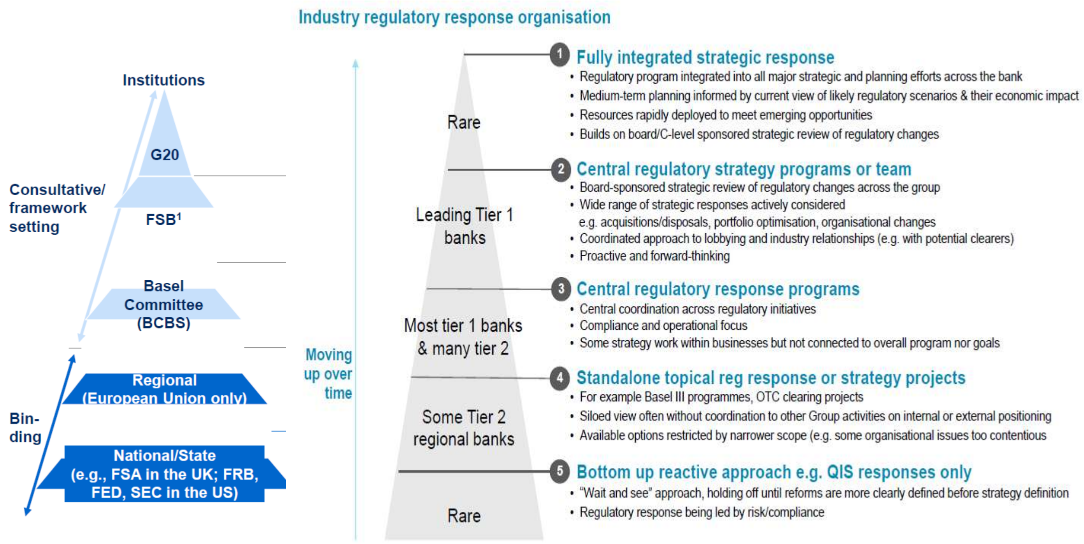


Figure 1.10: Left Panel: Regulation from a policy group view. Financial Stability Board was established in April 2009 as the successor to the Financial Stability Forum. *Source: McKinsey, 2012.* Right Panel: Industry regulatory response organization. *Source: Oliver Wyman, 2012.*

- Which initiative will indeed be **implemented** and to which extend? Whereas Basel III has a deadline some issues are still not finalized. Lobbying against the Dodd-Frank Act is that strong such that the outcome of the financial reform is likely to be reduced.

- **Complexity** vs. **effective** and **efficient** regulation. There different types of complexity issues:

- **Framework complexity**, i.e. the size, connectivity and implementability of a framework. The implementation of the Dodd-Frank Act requires that regulators create 243 rules, conduct 67 studies, and issue 22 periodic reports. The law itself consists of 2'300 pages. The final documents for the implications are likely to contain a factor of 5 to 10 times the 2'300 pages. If a nation intends to reshape the business model - here for the large U.S. banks - it is doubtful whether this is meaningful done in the form of the Dodd-Frank Act. First, the complexity of a systems generates also many loopholes. Second, complexity has at its best an ambiguous effect on the effectiveness. Third, despite the volume of the law it does not touches the key element of behavior or even behavioral changes of the bankers: they are forced to some business different, to stop some business, to search for alternative business ideas. But they are not forced to think about **values**. The complexity also shows if one consider the global structure of regulation, see Figure 1.12. The complexity of the structure raises the question about lack of leadership. Figure 1.10 shows that the pre crisis lack of leadership has changed to the present situation with G20 as the leading policy group and The Financial Stability Board (FSB) as the main coordination and standard setting body.[19] The BRIC (Brazil, Russia, India, China) states also are members of the FSB. This reduces the risk that important financial markets of the future are staying on completely the sideline in the regulation process. If we return to the Dodd Frank act, a fundamental question is to what extend are the U.S. able to handle large projects in the future due to the complexity of framework? This question is not only related to the regulation of the financial markets but also to other initiatives such as the Health Care Reform of the Obama administration.
- **Channel complexity**, i.e. the dependence between financial markets, governments, monetary policy, global macro economy, etc. The financial crisis 2008 showed that risk can disperse with a speed and in different economic and financial channels that make a global regulatory approach necessary. Figure 1.11 shows part of the dependencies related to the Greek-Eurozone crisis.

  The figure shows that the European Central Bank (ECB) can buy Greek bonds. This would guarantee that Greece will not pay to large credit spreads or even worse, the ECB guarantees a demand function for Greek bonds. ECB obtains the money to buy the bonds from the different central banks of the Euro zone. If this amounts become considerable for the individual central banks they will be recapitalized by the public budgets. The public also acts as a lender of last resort in other channels. First, private banks which are holding Greek bonds can run into problems if the bonds lose value and the asset value shrinks. Second the IMF can define rescue packages for Greece. But the IMF is funded by governments, i.e. public budget funding occurs also

[19] '...to coordinate at the international level the work of national financial authorities and international standard setting bodies and to develop and promote the implementation of effective regulatory, supervisory and other financial sector policies in the interest of financial stability.' FSB homepage.

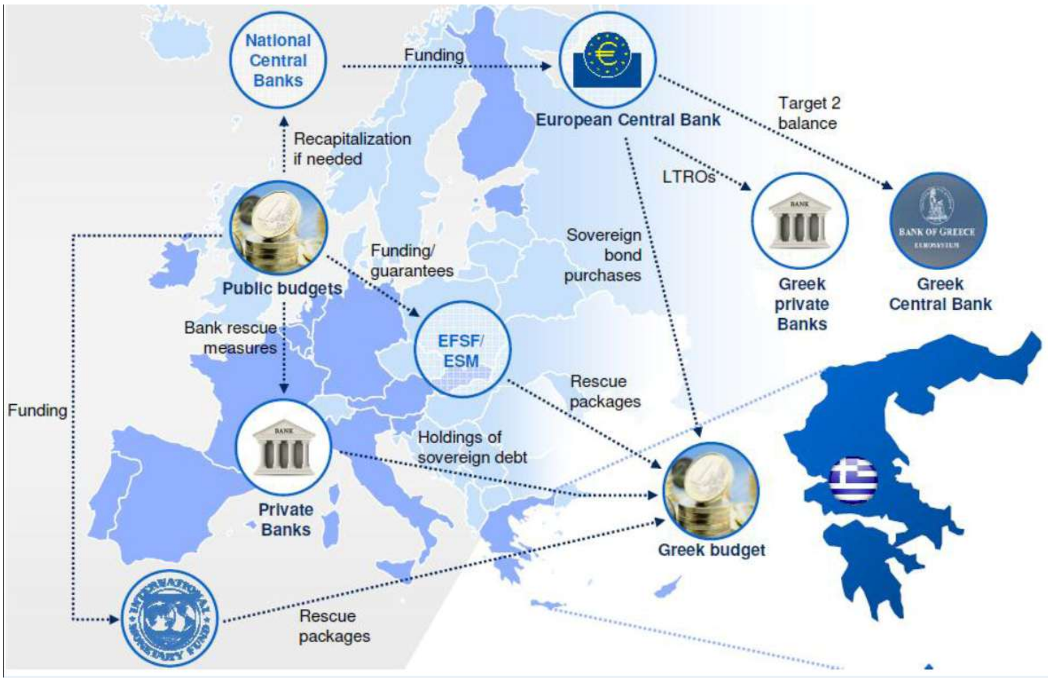


Figure 1.11: Representation of a systemic event hitting Greece and related financial channels. TARGET2 is an interbank payment system for the real-time processing of cross-border transfers throughout the European Union. LTROs are Long Term Refinancing Operations of the ECB.*Source: McKinsey (2012).*

in this channel. Finally, the European Financial Stability Facility (EFSF) is a special purpose vehicle financed by members of the Eurozone to address the European sovereign-debt crisis. As an example *'A Greek importer, for example, might place an order with a German company. Payments to and from the accounts of the buyer and seller are channeled via central banks, so the German exporter's bank gets a credit with the Bundesbank, which in turn has a claim on the ECB. The Greek importer's bank owes its local central bank, leaving the Bank of Greece with a debit at the ECB.' Source: J. Glover, 2012..*

- **Financial market complexity** describes the complexity of the web of link in financial markets such as the interbank market for example. We provide an overview over some questions in Section 1.2.4.

The different types of complexity impact different risks. Financial market complexity and channel complexity both impact the stability of financial system. The latter one is also a key factor for political stability. The key state variable to regulate financial market complexity is the regulation of the balance sheet since behavior, compensation and motivation of financial market participants strongly follows from the success of the financials. Mastering channel complexity is much more difficult since there is no clear cut target as the balance sheet for financial market regulation.

- **Competition** between financial centers vs. global regulation. Competition is to a large extend shaped by **inconsistencies in international regulation** between different regions.

  - US: Inconsistencies are a delayed introduction of Basel III, prohibition to hold commodity assets, FATCA (i.e. U.S. persons abroad are taxed, see discussion in the Tax Section), Volcker Rule and the Swap dealer registration are not consistent with regulations in Europe and APAC. Foreign entities will exit or scale back their onshore U.S. business.
  - Europe: Inconsistencies compared to other regions are the short selling bans, the introduction of financial transaction tax and the increasing political pressure on compensation. Compared to U.S. banks, European banks on average are overleveraged with capital and liquidity problems (US banks are cash rich), the banking sector is concentrated and the majority of corporate debt is retained on the bank balance sheet contrary to the U.S. where corpoartes use the capital markets to a much stronger extend to raise capital. This will lead to an asymmetric entrance of U.S. banks in Europe and given the U.S. inconsistency to a higher market share increase in commodities.
  - APAC (Asia and Pacific). Generally, the APAC states are lacking behind Europe and the U.S. in the implementation of new regulations and their approach is less stringent. This will lead to a relocation of trading business to Asia.

Summarizing, APAC will be the winning region, the U.S. will move sideways in the global competition and Europe will lose.

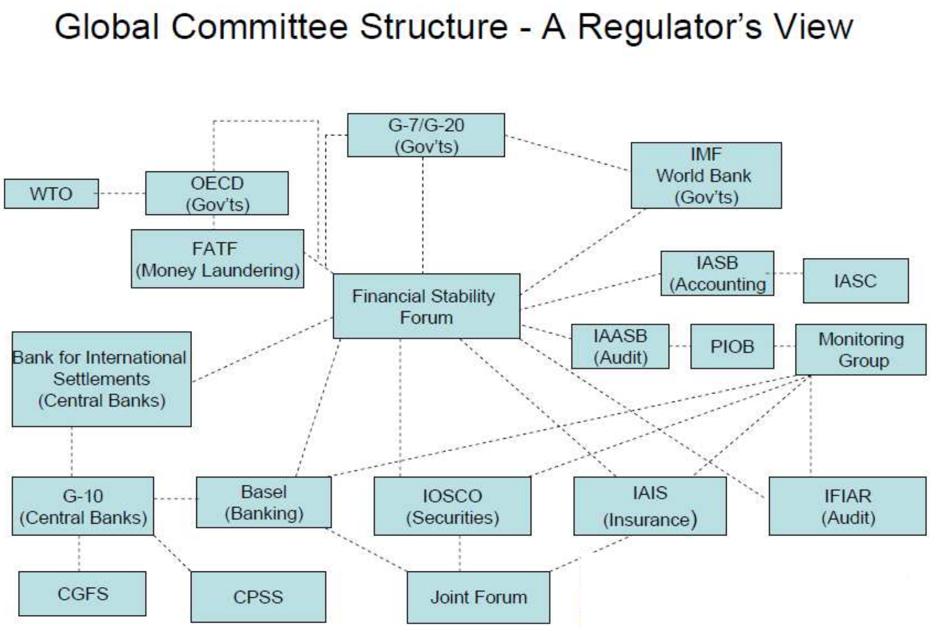


Figure 1.12: Topology of global regulation as of 2007. *Source: Sloan and Fitzpatrick (2007).*

1.2.4 First Approach to Systemic Risk

We consider systemic risk definition, - characteristics and some models types.

1.2.4.1 Empirical Studies

Most of the pre-crisis **empirical studies** on contagion in interbank networks are based on partial information on the bilateral exposures (aggregate balance sheet data). They estimate missing exposures with a Maximum Entropy method.[20] A part of the literature considered the network structure of the financial system in specific countries.[21] However,

[20] See Sheldon and Maurer (1998), Upper and Worms, (2004) Wells (2004), Elsinger et al. (2006) Degryse and Nguyen (2007) .

[21] See Furfine (2003) in the US, Upper and Worms (2004) in Germany, Agnes Lubloy (2006) in Hungary, van Lelyveld and Liedorp (2006) in the Netherlands, Wells (2004) and Elsinger et al. (2006a) in Austria, Wells (2004) in the UK, Mistrulli (2007) in Italy.

the Maximum Entropy method is found to underestimate the possibility of default contagion (Mistrulli (2007), van Lelyveld and Liedorp (2006)). Recent empirical studies are based on complete data sets. The study of Boss et al. 2004 for Austria and of Moussa et al. (2010, 2011) for Brazil. The findings of both studies are comparable. They have some strong implication for statistical model building of systemic risk. First, the use of famous Erdoes-Renyi, i.e. each edge is included in the graph with probability independent from every other edge or equivalently, all graphs with the same nodes and edges/links have equal probability to be drawn, leads to a Poisson node degree distribution. More precisely, consider a graph with n nodes and M edges. Each edge is included in the graph with probability p independent from every other edge. Equivalently, all graphs with n nodes and M edges have equal probability. For n to infinity with np = constant the degree has a Poisson distribution:

$$P(\text{degree} = k) = \frac{(np)^k e^{-np}}{k!} \ .$$

The graph has no heavy tails. The analysis of the Brazilian and the Austrian network leads to more heavy-tailed node degree distribution, i.e. both a power law distributions[22] follows for the in-degree (the number of debtors in the nodes) and the out-degree (the number of creditors) and a Pareto distribution for the exposures. The Erdoes-Renyi networks also shows a too low clustering compared to the financial systems. Besides this statistical arguments the fact that in an Erdoes-Renyi model links are formed independent from any other link violates basic behavioral aspects in banking: Bank's decision to whom they connect for business is not a random process, neither in the formation of a link and also not in the possible deletion of links in cases where interbank markets are under stress. In summary, the Erdoes-Renyi model

- shows too low clustering,
- fails to produce the heavy-tailed node degree distribution.

The model of Minca in the next section allows for networks which generate these empirical findings, i.e. he show the importance of taking into account the heterogeneity of financial networks when discussing issues of financial stability and contagion. Figure 1.13 illustrates the topology of the UK and Brazil banking system. We stress that it is important to describe what the links and nodes actually mean.

The **financial system** showed in the **last decade** three remarkable developments, see Haldane (2009): The **increase in complexity, trading speed and homogeneity**. The increase in complexity has several dimensions. First, the scale and interconnectivity of the international financial network has increased significantly over the past two decades. Haldane (2009) shows that the size of nodes if the network of countries is considered increased by a factor 14 and links have become both fatter and more

[22] A system exhibits a power or Zipf Law if the tails of the distributions exhibit a linear decay in log-scale.

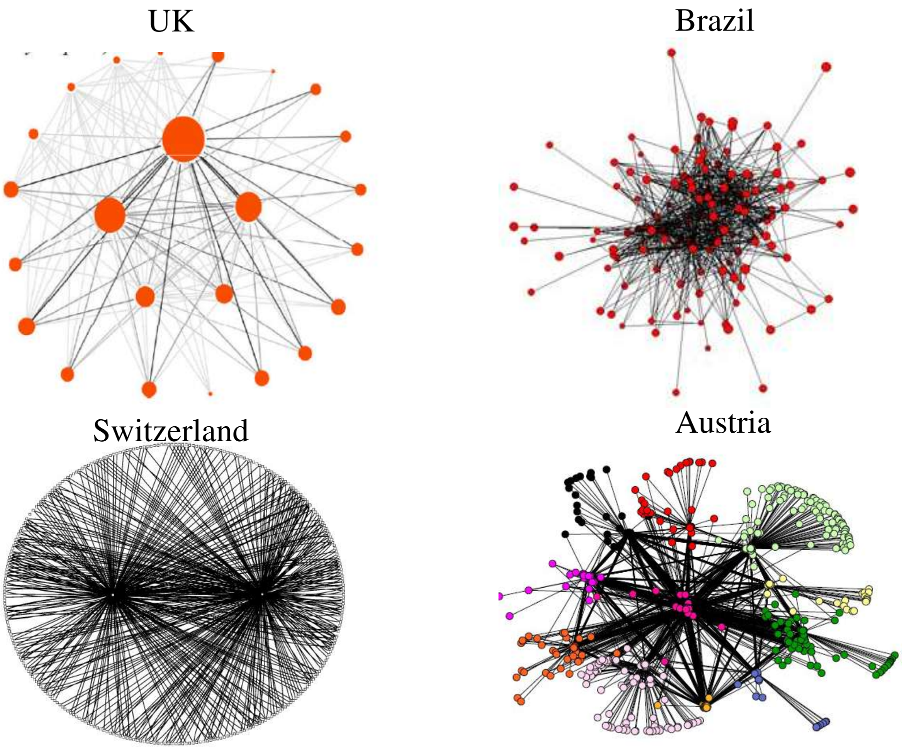


Figure 1.13: UK: UK interbank market Q1 2008 of large exposures. A large exposure is one that exceeds 10% of a lending bank's eligible capital during a period. Eligible capital is defined as Tier 1 plus Tier 2 capital, minus regulatory deductions. Each node represents a bank in the United Kingdom. The size of each node is scaled in proportion to the sum of (1) the total value of exposures to a bank, and (2) the total value of exposures of the bank to others in the network. The thickness of a line is proportionate to the value of a single bilateral exposure. *Source: Bank of England (2009).* Austria: scale-free structure *Source: Boss et al., (2004).*, Switzerland: sparse and centralized structure. *Source: Mueller, (2004).* Brazilian interbank network, December 2007. A scale-free structure. The number of financial conglomerates is n = 125 and the number of links in this representation at any date does not exceed 1200. The network has a heterogeneous and complex structure, some highly connected institutions playing the role of hubs while others are at the periphery *Source: Cont et al. (2010).*

frequent by a factor of 6 between 1985 and 2005. But not only the topology became more complex also the information propagating in the network faced increasing complexity. This is largely due to the industry approach of slicing, re-bundling and redistributing risk, i.e. the structured finance approach. This wide spread risk transfer technique lead to basic and enormous sized credit derivative markets such as the CDS market and also to the complex structured finance securities such as CDO, CDO-squared and so on. This risk transfer technique together with the 'originate and distribute' business strategy in the US led to *long value chains* in the network. The increase in **homogeneity** is due to the follow-the-leader approach by the banking firms. Each new trend is adopted by the whole industry, starting from leveraged loan business they switched to structured finance (CDO) and all institutions started to hold risk in form of such complex assets instead of taking the role of risk transformator. Not only the same business strategy was used but also risk loading to the need to leverage the own balance sheet was standard. This search for new assets led the institutions the way down from top rated and relatively simple assets to assets of a lower creditworthiness and/or a higher complexity.

1.2.4.2 Risk Description: Probabilistic Network Modelling

A network consists of nodes and links. The **nodes** represent banks and each **directed link** reflects a business relation between two nodes. Contrary to physical or biological systems, financial system nodes are more **complex**. They are high dimensional objects - the financial dimension given by the balance sheet and the income statement as well as other dimensions such as intangible assets and reputation characterize a node. This complexity of the nodes allows for different channels of risk migration in banking networks, see Figure 1.14.

We consider the different channels of risk migration in a network.

First, there are directed links representing **cash flows** such as margin calls or funding cash flows. These links represent **liquidity risk**. The initial illiquidity and default on payments may transmit to counter parties which may become also illiquid - a **cascade of illiquidity** follows. Such margin runs can arise from margin payments on outstanding derivatives due to large jumps in the mark-to-market values of the derivatives. Credit default swaps are particularly prone to large jumps, even in absence of default of the reference entity.

Second, **balance sheet exposures**, i.e. the expected loss on outstanding claims in case of counter party default, are represented by directed links. Every interbank liability is another bank's asset. Hence unsecured interbank assets are endogenously determined by the network links. This defines the **insolvency channel**: A bank becomes insolvent (default) if its capital buffer cannot withstand the losses due to direct exposures. The liquidation of an illiquid portfolio of a defaulted bank on the market has a price feedback effect on the portfolios of other banks holding similar assets. They in turn may become insolvent too leading to a potential **insolvency cascade**. To take into account different markets and different links between assets and liabilities one needs to introduce a fine

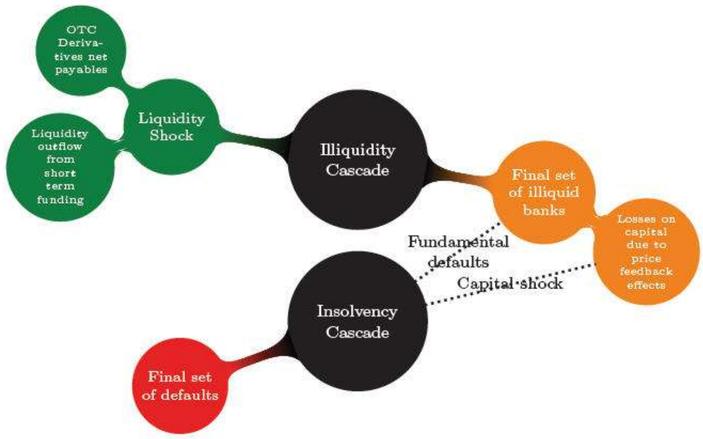


Figure 1.14: Different risk channels. *Source: Minca, 2011*

structure on both the asset and liability side, i.e. one has to distinguish between secured and unsecured securities, liquid and illiquid assets.

Finally, a third kind of contagion is due to **similarity links**: Two banks are linked in this way if they possess the same assets where the asset possession does not induce any contractual obligation. This type of link has possible price spillover effects: Bank A needs to sell the assets in a fire sale way where bank B owes the same assets. This is the **capital shock channel**.

The mechanics of distress propagation in financial network through channels with a financial measurable unit can be introduced as follows.[23] Let e_{ij} represent any kind of exposure, i.e. the maximum loss related to direct claims, of bank i to j. If $e_{ij} < 0$, then j has a liability to i, see Figure 1.15. By the limited liability rule the capital E_i before a shock of size ϵ_i becomes

$$\tilde{E}_i = \max(E_i - \epsilon_i, 0) \ .$$

The set of default sequences D_k is defined as follows. D_0 is the set which contains all banks which are in a default status at initiation. If bank j defaults in the next step, i faces a loss e_{ij}. Then,

$$D_1(e, E) = \{\text{banks } i | \ E_i < \sum_{j \in D_0} e_{ij}\}$$

[23]The model is based on the thesis of A. C. Minca, (2011).

is the next set of defaulted banks. Therefore, D_1 consists of all initial defaulted banks D_0 plus all banks, where the capital puffer turns out to be low due to the losses of the D_0-banks. **The insolvency cascade** is represented by iterating the above procedure which ends with the last possible default, i.e. we have in step k:

$$D_k(e, E) = \{\text{banks } i | \ E_i < \sum_{j \in D_{k-1}} e_{ij}\} \ , \tag{1.1}$$

all banks are considered such that their capital is smaller than the losses from connected banks defaulted in step $k-1$. One then considers a sequence (e_n, E_n) of financial networks, indexed by the nodes n and the in- and out-node degrees. We set G_n to be equal to the set of all weighted directed graphs which match the sequence of financial networks. The goal is then to study the behavior of the fraction of defaults

$$a_n(G_n, E_n) = \frac{|D_n(G_{n-1}, E_{n-1})|}{n}$$

for n to infinity (analytical) or numerically for finite n.

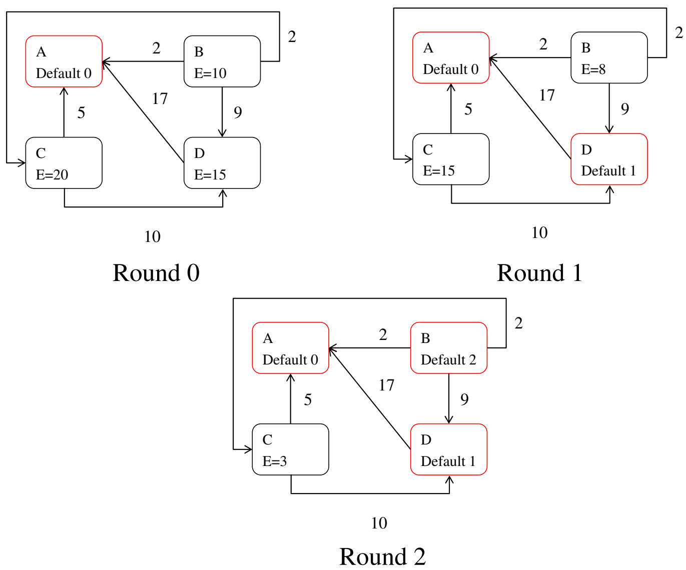


Figure 1.15: A toy network which illustrates a solvency cascade. The links represent exposures net of collateral and in the nodes the capital is shown.

In a similar way an **illiquidity cascade** is defined. The capital condition is replaced by a liquidity condition in the D-sequence. To achieve this the assets are decomposed into liquid and illiquid ones and in- and out-cash flows between counter parties are introduced. These cash flows represent for example maturing loans, coupon payments, margin calls from OTC derivatives. Similar to the insolvency cascade a set of illiquid banks D_k follows.

We consider as an example the model of Gai and Kapadia (2007). They consider n banks with directed links: Interbank loans define links. Links are directed. The authors use the following terminology and make the following assumptions:

- Incoming links = interbank assets.
- Outgoing links = interbank liabilities.
- In-degree: number of links that point into the node.
- Out-degree: number of links that point out of a node.
- They assume that: Average in-degree = average out-degree.
- Total liabilities of each bank are normalised to unity.
- The total interbank liability position of every bank is evenly distributed over each of its outgoing links and independent of the number of links the bank has.

The distribution of the in-degree and out-degree are unspecified. The joint distribution of in- and out-degree governs the potential for the spread of shocks through the network. Degree distribution is arbitrary, i.e. the authors do not assume a particular topology. This counteracts the critique many prior studis faced of using artifical or non realistic network structures. They split the assets split in interbank assets A^{IB} and illiquid assets A^M. They assume that the total interbank asset position A^{IB} is evenly distributed among the banks independent of the number of links: This mimics the diversification motivation. Solvency for bank i means

$$(1-\phi)A_i^{IB} + qA^M - L_i^{IB} - D_i > 0$$

with D_i the deposits, ϕ the fraction of banks with obligations to bank i that have defaulted and q the resale price of the illiquid asset. We make a zero recovery assumption. Using the capital buffer E_i the solvency condition reads:

$$\phi < \frac{E_i - (1-q)A_i^M}{A_i^{IB}} \ .$$

Assume that at time 0 all banks are solvent and that at time 1 a single back i faces a shock. Let j_i be the number of in-links for i. Since linked banks each lose a fraction $1/j_i$ of their interbank assets when a single counter party defaults, the only way default can spread is if there is a neighboring bank for which

$$\frac{E_i - (1-q)A_i^M}{A_i^{IB}} < \frac{1}{j_i} \ .$$

Hence, $D_0 = \{i| \frac{E_i-(1-q)A_i^M}{A_i^{IB}} < \frac{1}{j_i}\}$. Banks that are exposed in this sense to the default of a single neighbor are called **vulnerable** and other banks as safe. Vulnerability of a bank depends on its in-degree. Let v_j be the probability that a bank with in-degree j is vulnerable and let p_{jk} be the joint degree distribution of in-degree j and out-degree k. Why is this joint distribution important? Contagion spreads via incoming links. For contagion to spread beyond vulnerable first neighbours, those neighbours must have incoming links from other vulnerable banks. Going via incoming links, we are interested in second neighbours of vulnerable banks. The degree distribution of second order neighbours r_{jk} is proportional to kp_{jk}. This is a result due to Feld (1991), see also Newman (2003). Bank which are not vulenerable to liquidity hoarding behavior of their neighbours are called **safe**. The quantities (v, p) are the basic objects to be analyzed: In large networks, for contagion to spread beyond the first neighbors of the initially defaulting bank, those neighbors must themselves have outgoing links to other vulnerable banks. To obtain a condition for the transmission of **shocks** one works with the probability-generating function for the joint degree distribution of a vulnerable bank

$$G(x, y) = \sum_{j,k} v_j p_{jk} x^j y^k \ .$$

The information content of this function is the same as for the degree distribution and the vulnerability distribution. But it allows us to work with sums of independent draws from different probability distributions and it generates all the moments of the degree distribution of only those banks that are vulnerable. Since every interbank asset of one bank is an interbank liability of another, the average in-degree $z := \sum_{j,k} p_{jk}$ in the network must equal the average out-degree. The single-argument generating function $G_0(y) = G(1, y)$ can be used for the number of links leaving a randomly chosen vulnerable bank. $G_0(1) = \sum_{j,k} v_j p_{jk}$ yields the fraction of banks that are vulnerable. The function $G_1(y)$ defines the number of links leaving a bank reached by following a randomly chosen incoming link. It is given by

$$G_1(y) = \sum_{j,k} v_j r_{jk} y^k$$

where $v_j r_{jk}$ is the degree distribution of a vulnerable bank that is a random neighbor of our initially chosen bank. Finally, if $H_1(y)$ is the generating function for the probability of reaching an outgoing vulnerable cluster of given size (in terms of numbers of vulnerable banks) by following a random outgoing link from a vulnerable bank we obtain the recursive equation

$$H_1(y) = 1 - G_1(1) + yG_1(H_1(y)) \ .$$

That is, the total probability of all possible forms equals the sum of probabilities of hitting a safe bank $(1 - G_1(1))$, hitting only a single vulnerable bank, hitting a single vulnerable bank connected to one other cluster, two other clusters and so on. This sequence is equal to

$$yG_1(H_1(y)) = y \sum_{j,k} v_j r_{jk} (H_1(y))^k \ .$$

It is not possible to find a closed-form expression for the complete distribution of cluster sizes but numerical analysis can be used. Using a Poisson model, the authors show using simulations: *'We find that financial systems exhibit a robust-yet-fragile tendency: while the probability of contagion may be low, the effects can be extremely widespread when problems occur. The model also highlights how a priori indistinguishable shocks can have very different consequences for the financial system, depending on the particular point in the network structure that the shock hits. This cautions against assuming that past resilience to a particular shock will continue to apply to future shocks of a similar magnitude. And it explains why the evidence of the resilience of the financial system to fairly large shocks prior to 2007 (e.g. 9/11, the Dotcom crash and the collapse of Amaranth to name a few) was not a reliable guide to its future robustness. The intuition underpinning these results is straightforward. In a highly connected system, the counter party losses of a failing institution can be more widely dispersed to, and absorbed by, other entities. So, increased connectivity and risk sharing may lower the probability of contagious default. But, conditional on the failure of one institution triggering contagious defaults, a high number of financial linkages also increases the potential for contagion to spread more widely. In particular, high connectivity increases the chances that institutions that survive the effects of the initial default will be exposed to more than one defaulting counter party after the first round of contagion, thus making them vulnerable to a second-round default. The effects of any crises that do occur can, therefore, be extremely widespread.'*

1.2.5 Balance Sheet Regulation - Basel III

The **Basel III** accord requires **more and better capital** from the banks compared to Basel II as a buffer against solvency risk, it gives equal weight to **liquidity risk** regulation and the requirements for market and credit counter party risk are increased: Contrary to Basel II not only capital but the whole balance sheet is regulated. To reduce the impact of model risk in the calculation of the risk weighted assets, which matter for the capital charge, a model independent **maximum leverage** of 3 percent is introduced for systemic relevant banks by 2018. This means that the asset side length can be no longer 33 times the value of capital. than On the liquidity sides two measures are introduced. One reduces the probability of a liquidity stress (going concern) and the other one is used to increase the survival time once an institution is in a liquidity stress. The timeline for the implementation starts 2013 and ends 2019. More precisely the Committee - the Basel Capital Accord Committee - proposes a series of measures. These measures can act on the state variable capital (smoothing and buffer stock formation) or decision variables as provisions which can forced to be more forward looking. We have (McKinsey (2012), Deloitte (2012)):

- Changes affecting the **asset** side:
  - New market risk and securitization framework: Introduction of a stressed value- at-risk (VaR) capital requirement and an increase of capital requirements for re-securitizations in both the banking and the trading book.

  - Counter party credit risk: Introduction of additional charges for counter party credit exposures arising from banks' derivatives, repo, and securities financing activities (see CVA).
  - Higher Risk Weighted Assets (RWA) for financial intermediary (FI) exposures: About 30 percent higher RWA for FI exposures, for example interbank exposures.
  - **Liquidity coverage ratio (LCR)**: Introduction of a global minimum short-term 30-days liquidity standard, requiring banks to reach sufficient amount of holdings of highly liquid assets. LCR is defined as:

    $$\frac{\text{High Quality Liquid Assets}}{\text{Net Cash Outflows in 30d Period}} > 100\% \ .$$

    The 30d period considered is a **stress event**: One assumes that the bank's debt is reduced by three notches using one of the official ratings, wholesale funding is complete lost, a fraction of deposits is lost, collatal posted is valued less (equivalently, the haircuts are raised, i.e. the bank has to deliver more collateral) and the lines of credit are drawed down. LCR will be monitored by both the Basel Committee and the supervisor in order to officially make it mandatory with effect from 1 January 2015.

- Changes affecting the **liability** side:

  - **Net stable funding ratio (NSFR)**: Introduction of a long-term structural ratio to address liquidity mismatches. This ratio is defined as:

    $$\frac{\text{Amount of Stable Funding}}{\text{Required Amount of Stable Funding}} > 100\% \ .$$

    The categories are defined as follows. The regulator defines categories of funding for the numerator. Each funding is then attributed to a category and multiplied by a category specific factor; the **available stable funding factor** (ASF). Figure 1.18 shows the categories and the factors. NSFR will be introduced as a minimum standard as from Jan 2018.

- Changes affecting **capital**:

  - **Minimum (core) tier 1 ratio**: Increase of the minimum common equity (CET1, i.e. shareholders equity and retained earnings) requirement from 2 percent to 4.5 percent. The capital ratio is defined by

    $$\frac{\text{Regulatory Capital}}{\text{RWA}} > x\%,$$

    i.e. the regulatory capital has to exceed x percent of the risk weighted assets. The figure x is discussed below. The national implementation will start on 1 January 2013. From Jan 2013 onwards, banks will have to meet the following minimum capital requirements expressed in risk-weighted assets: 3.5% share capital, 4.5% Tier-1 capital and 8% total capital. During the transitional period from Jan 2013 up to and including 2019, these ratios will gradually be stepped up to 4.5% share capital, 6% Tier-1 capital and 8% total capital.

  – **Conservation buffer**: Introduction of an additional (counter-cyclical) capital conservation buffer between 0 % and 2.5 % to withstand periods of stress. The buffer will be build up along gradual lines to a percentage of 2.5% from Jan 2016 through Jan 2019. Thus, banks will ultimately have to hold 10.5% of their total capital expressed in risk-weighted assets.
  – **Capital quality:** Requirement to form tier 1 capital predominantly through common shares and retained earnings.
  – **Capital deductions**: International harmonization of capital deductions and prudential filters. National supervisors will gradually introduce additional allowable deductions from bank capital such as deferred tax assets and investments in financial institutions from Jan 2014 - Jan 2018.
  – SIFI capital surcharges: Higher capital charges for systematically important financial institutions (SIFIs).

- Changes affecting the **whole** balance sheet:
  – **Leverage ratio**: Introduction of a non-risk-based measure of capital structure. The period from Jan 2013 through Jan 2017 will be a parallel run period. During this period, the development of the Leverage Ratio will be monitored. The intention is to migrate the Leverage Ratio into the Pillar 1 requirements as from Jan 2018.
  – Supervisory review of new **remuneration** policies: Closer review of remuneration policies, especially in case of weakening capital buffers.

For to big-to-fail banks, i.e. the SIFIs, regulation will possibly introduce a **counter cyclical buffer (CCB)**. The counter cyclical buffer is different than the conservation buffer which also shows some counter cyclical behavior. The goal of the CCB is increase the resilence of the banking system and to act as a damping factor in times of excessive credit lending (so called leaning against the wind): The objective of counter cyclical capital standards is to encourage banks to build up buffers in good times that can be draw down in bad ones. Buffers should not be understood as the prudential minimum capital requirement. They are unencumbered capital in excess of that minimum, so that capital is available to absorb losses in bad times. Counter cyclical capital buffer schemes can be thought of as having the objective to limit the risk of large-scale strains in the banking system by strengthening its resilience against shocks. As a model consider

$$\text{Ratio}_t = \frac{\text{Credit}_t}{\text{GDP}_t} \times 100\%.$$

Using a statistical filter, the Hodrick-Prescott filter for example, one extracts from the Ratio the Trend at a given date. This defines the gap

$$\text{Gap}_t = \text{Ratio}_t - \text{Trend}_t \ .$$

The buffer add-on of the CCB is then defined as a piecewise linear function. The add-on is zero if the gap is below a given floor, the add-on increases linearly up to a maximum level for the gap. Figure **??** illustrates the impact for the UK.

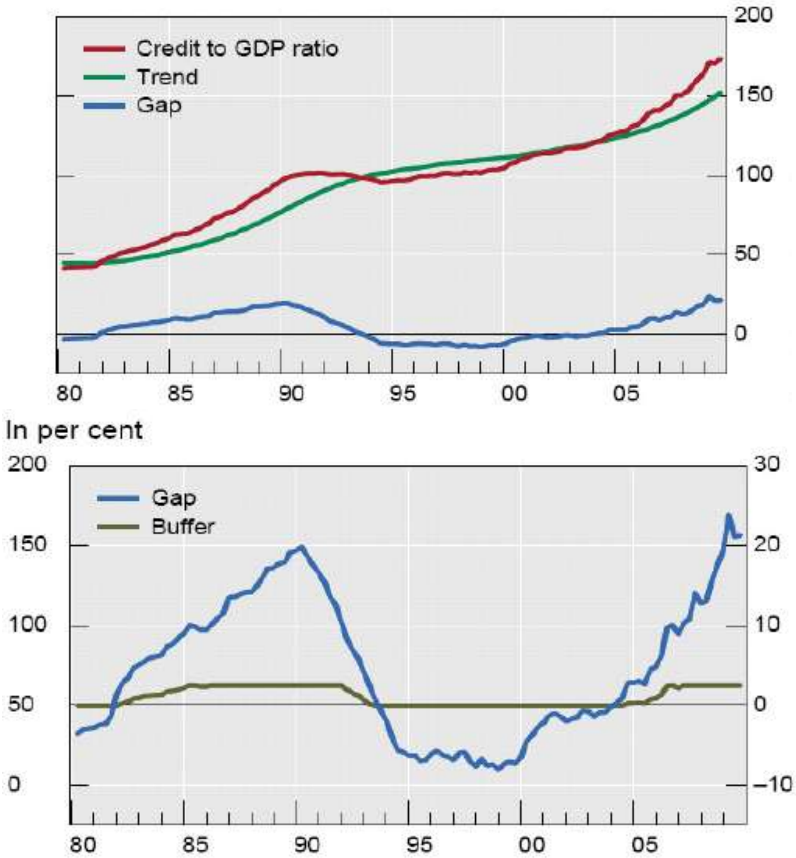


Figure 1.16: Counter cyclical buffer for the UK. The upper panel shows the evolution of the credit to GDP ratio, the trend component and the gap. The figure shows the boom periods in the 80s of last century and in the second half of the last decade up to the financial crisis. The lower panel shows the relation between the gap and the capital buffer. The buffer is largest during the boom period, where it even is often capped at its maximum level, and low or zero in periods of weak or negative economic growth.*Source: BIS, UK national data, BCBS.*

The methodology how regulatory capital is calculated follows the Basel II approach. But some models are changed, new models are introduced or some model free regulation (maximum leverage) is used in Basel III. First, for the different risk factors market risk, credit risk and operational risk different approaches exist to calculate the **Risk Weighted Assets (RWA)**. Some are not risk sensitive, i.e. they are not derived from a statistical model. They are applied by smaller banks. Statistical models are used by larger banks. This means that larger banks can either develop own models (advanced approach for operational risk) or use internal models to generate input parameters in predefined statistical models (advanced models for credit risk). We note that it is at the national regulatory authority discretion to decide whether a bank can use a more sophisticated model or whether the bank has to. Further, internal models need to be approved by the regulatory body. If one compares a mortgage portfolio and applies to the portfolio the standardized approach (which is a non-statistical model) and then uses an advanced credit risk model for the same portfolio one typically observes:

- The RWA are lower using the advanced model if the portfolio has a good credit risk quality compared to the standardized model.
- The opposite is true if the credit worthiness of the counter parties in the portfolio is low.

This raises the issue of **regulatory aribtrage**. A bank would like to choose an advanced model in cases where the RWA are lower compared to a non-statistical model and vice versa. But there is no such cherry picking: A bank which wants to use an advanced credit risk model has to use this theoretically for all counter parties: For the mortgage portfolio of retail clients, for the loans to corporate clients, for credit risk in the interbank market relations, for counter party risk in the derivative transactions, etc. In practice, a 100 percent fullfillment is not possible: To develop a statistical model for credit risk in start up financing is meaningless since there are no data to calibrate the parameter, to backtest the model and to see how the model would have behaved in a stress scenario - these are all requirements a bank has to fullfill in order to apply an advanced model. Therefore, the regulator will allow to treat some parts of the portolio which are not relevant from a solvency risk view using simpler, non-technical approaches. Finally, if a bank fails to meet the standards for a part of the portfolio the regulator will use multiplier in the calculation of the RWA.

We discuss one particular model for credit risk below. The logic of the capital requirement is that

$$\frac{\text{Regulatory Capital}}{\text{RWA}} > x\%.$$

Basel III restricts the parts of capital which classify as regulatory capital: Tier 1 are common shares and retained earnings, Tier 2 is harmonized and Tier 3 is no longer allowable. All other equal the numerator decreases for an institutions which requires to raise more Tier 1 capital or to reduce risks in the RWA calculation. Second, the risk coverage will be strengthened. For example the capital requirements for counter party credit exposures arising from banks' derivatives transactions are strengthened or capital incentives are introduced to move OTC derivative contracts to central counter parties. This might again affect the capital level, the OTC example, or the RWA increase, see the Section Credit Risk for details. Finally, the $x\%$ which are under Basel II 8% increase over time. This will have the effect that too risky business, i.e. business which leads to high additional capital and/or high RWA will be no longer profitable. The capital accord following Basel III and a national finalization, here the Swiss Finish for the to largest banks UBS and Credit Suisse, are shown in Figure **??**.

While the industry raised up to 80 percent of the required capital there is still much potential to reduce the RWA (Oliver Wyman 2012). RWA can be decreased either by technical mitigation of risks or strategic exits. Since currently in Europe only 20 % of European mid-size companies have direct access to capital markets compared to the U.S. (80 percent), it is expected that banks will first reprice their loans due to an increased capital charge and offer European corporates the possibility to issue debt for financing. Since this shift affects mid-size corporates this risk mitigation off the bank's balance sheet can be favorable for smaller, regional banks which already have long-term lending relationship and have access to local bond markets.

Given the long period of implementation of Basel III, the existence of Basel 2.5 [24] there will be **confusions** when capital adequacy is calculated. Confusion increases if other capital charge definitions are used. As an example, we consider the **Financial Stability Report** of the Swiss National Bank. The role of this report is:

In its Financial Stability Report, the Swiss National Bank presents its assessment of the Swiss banking sector's and financial market infrastructures' stability. The Bank publishes this report as part of its contribution to the stability of the financial system – a task assigned to it under the new National Bank Act. In the Financial Stability Report, the National Bank focuses on trends that are observable at the levels of the banking system, the financial market infrastructures, the financial markets and the macroeconomic environment. The main purpose of the report for the National Bank is to draw attention to strains or imbalances which could pose a threat to system stability in the short or the longer term. The Bank thus tracks developments in the banking sector from a macroprudential perspective. This task supplements that assigned to the Swiss Financial Market Supervisory Authority (FINMA), which is responsible for supervision of the banks at the level of the individual institutions (microprudential supervision). In addition, the Financial Stability Report provides information about its activities regarding the oversight of financial market infrastructures. Source: Swiss National Bank, homepage.

[24] *'Basel III builds upon and enhances the regulatory framework adopted by Basel II and Basel 2.5, which now form integral parts of the Basel III framework: Basel II improved the measurement of credit risk and included capture of operational risk, was released in 2004 and was due to be implemented from year-end 2006. Basel 2.5, agreed in July 2009, enhanced the measurements of risks related to securitization and trading book exposures.2 Basel 2.5 was due to be implemented no later than 31 December 2011.'* Source: BIS, 2012.

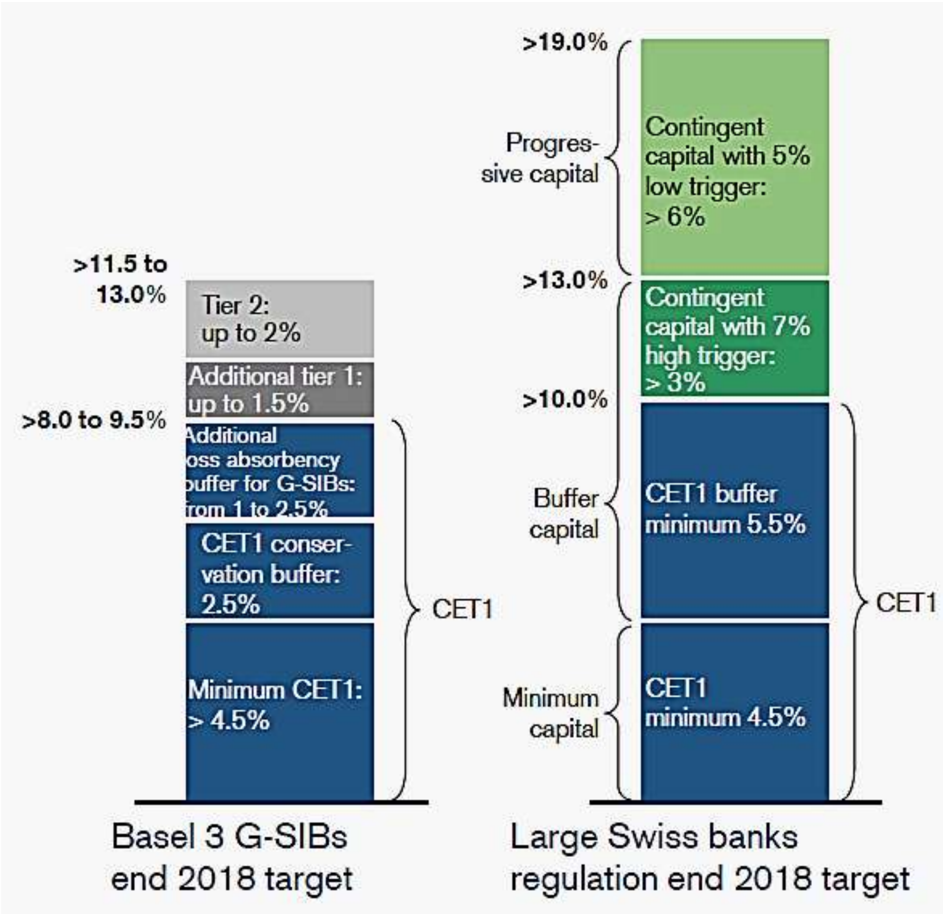


Figure 1.17: Left Panel: Capital requirements for global systemically important banks (G.SIBs). An additional 1% surcharge in the additional loss absorbency buffer might be applied to provide a disincentive for banks to increase materially their global systemic importance in the future. after Basel III. Right Panel: Capital accord (Swiss Finish to Basel III) for the two large Swiss Banks UBS and Credit Suisse (status January 2012). Swiss Expert Commission requiring CET1 ratio of 10%, compared to a 8% to 9.5% ratio under Basel III. Additional buffer and progressive component capital to achieve a Total Capital ratio of 19% . A requirement which can be met with a combination of 7% high trigger and 5% low trigger contingent capital instruments. As Tier 1 and Tier 2 instruments under Basel III will require loss absorbency at the point of non-viability, they will need to include contingent capital like features.*Source: BIS, Credit Suisse Group.*

The report in 2012 then reports that the loss-absorbing capital of the two Swiss SIFIs UBS and Credit Suisse below the level needed to ensure sufficient resilience. The National Bank calculated the loss-absorbing figures by using all definitions of Basel III but the low trigger contingent capital, see Figure **??**. The reason is that
'... These are mainly intended for the Swiss emergency plan and the restructuring or wind-down of the remaining bank units, and are therefore not considered in this 'going concern' perspective. Source: Swiss National Bank'.

Although the two banks had at the end of first quarter 2012 a capital ratio following Basel III of 5.9 and 7.5 percent the loss-absorbing capital was only 1.7 and 2.7 percent respectively. The announcements of these figure put in particular Credit Suisse under pressure - both in raising new regulatory capital at a faster pace than planned and by a drop of the share price. Besides these actions, there were a lot of confusions since bankers were pointing to the good-lucking Basel III figures but central bankers focused on the lower loss-absorbing numbers.

The increased cost of capital may boost **shadow banking** . Strictly speaking, there is no definition what the shadow banking area is since either an entity is a bank and regulated or it is not a bank. One considers often Credit Hedge Funds, Special Purpose Vehicles and Money Market Mutual Funds to be part of the shadow banking sector. There is no consensus about other entities such as Credit Card Institutions, Hedge Funds, Independent Asset Managers, Commodity Traders. The FSB (2011) released some recommendations for the shadow banking sector to mitigate systemic risk and regulatory arbitrage. Basically the recommendations increase the costs of capital and liquidity for banks when they reallocate resource to the shadow banking sector and they increase the costs of operations such as increasing transparency and information requirements for the end investors.

We consider the impact of some measures in more detail.

Before the financial crisis, regulatory capital was allowed to be smaller under the then active Basel II regime than it will be under Basel III, see Figure 1.17. This led to high leverage, i.e the ratio between assets and capital was large. Some investment banks faced leverage ratios between 50 and 100. I.e. banks had only down to 1 percent risk buffer. Among others, this leads to a high **return on equity (RoE)** value which in some banks has a direct impact on compensation and the dividend policy. To show this consider two dates 0 and T. The balance sheet reads at each date $A = L + E$ with A the assets, L the liabilities and E equity or capital. The Return on Assets (RoA) and the Return on Equity (RoE) are defined by[25]

$$\text{RoA} = \frac{A_T}{A_0} - 1 \ , \ \text{RoE} = \frac{E_T}{E_0} - 1 \ .$$

[25] We follow Beneplanc and Rochet (2010).

$\lambda = \frac{A}{E} > 1$ defines the **leverage factor** and $R = \frac{L_T}{L_0} - 1$ interest income on the liabilities. We get [26]

$$\text{RoE} = R + \lambda(\text{RoA} - R) \ . \tag{1.2}$$

Excess return on capital is proportional to excess return on assets multiplied by the leverage factor. This is also true if we replace the quantities by expected values. Let μ_X be the expected value of a random variable X. We get:

$$\mu_E = R + \lambda(\mu_A - R) \ .$$

Similarly for risk measured by volatility σ:

$$\sigma_E = \lambda \sigma_A \ . \tag{1.3}$$

Hence a higher leverage factor increases both RoE and risk. The latter one increases solvency risk, i.e. the risk that the asset value falls below the liability value ($P(A < L)$). The stronger A varies, the higher the risk that E vanishes. For a RoA of 8 percent, $R = 3\%$ and a leverage factor of 4 the expected return is

$$\mu_E = 3 + 4(8 - 3) = 23\% \ .$$

If the risk of the asset return is $\sigma_A = 10$, volatility of the capital increases to $\sigma_E = 4 \times 10\%$. Summarizing, an increasing leverage factor increases **jointly** RoE and insolvency risk. But the story does not end here. So far, **financing risk** or the liability side of a balance sheet was not considered. We analyze the relationship between interest rate risk management of the balance sheet and leverage. We assume that both assets and liabilities depend on a single interest rate r. For small changes in the rate r uniformly for all maturities **Macauley duration** D is a first order risk management figure. It is equal to the price weighted average time of receipt of cash flow. It is the time to maturity $D = T$ for a zero bond[27] such that this bond has the same interest rate[28] risk than a given portfolio of bonds. The small change in interest rate is given by the first derivative ($\frac{dp(t,T)}{dr}$) leads for a zero bond price $p(t, T)$ to

$$\frac{dp(t,T)}{dr} = -\frac{T}{1+r} p(t,T) =: -D^* p(t,T)$$

with D^* the modified Macauley duration. The percentage price sensitivity $\frac{dp(t,T)}{dr}/p(t,T)$ is proportional to the duration $T = D$ of the zero bond. Since duration is given by the

[26] To prove this rewrite the RoE:

$$\text{RoE} = \frac{E_T}{E_0} - 1 = \frac{A_T - L_T}{E_0} - 1 = \frac{A_T}{A_0}\frac{A_0}{E_0} - \frac{L_T}{L_0}\frac{L_0}{E_0} - 1 \ .$$

Inserting the definition of leverage and interest income at time 0 proves the claim.

[27] A zero coupon bond (or zero) is a bond which pays no coupons, has price 1 at maturity T and price $p(t, T)$ at any earlier time.

[28] More precisely, interest rate risk using the duration concept is defined as the risk if the term structure is shifted parallel by a small amount.

first derivative, the duration of a portfolio of bonds is equal to sum of weighted durations of the bonds. We construct synthetic bonds for the assets, liabilities and equity such that these bonds have the same duration as the assets, liabilities and equity. Since these three quantities are related by the balance sheet, a relation between their duration measures follows. To derive this, we note that from $E(r) = A(r) - L(r)$ we get $\frac{E'}{E} = \frac{A'-L'}{E}$ where the prime denotes differentiation w.r.t. r. This is equivalent to

$$\frac{E'}{E} = \frac{A'}{A}\frac{A}{E} - \frac{L'}{L}\frac{L}{E} \ .$$

Using $A/E = \lambda$, $L/E = \lambda - 1$ and defining the duration D_X of an asset X by $D_X = -(1+r)\frac{X'}{X}$ we get

$$D_E = \lambda D_A - (1-\lambda) D_L \ . \tag{1.4}$$

The sensitivity of a banks equity has two sources. First, risk transformation. That is typically the duration of the assets is longer than that one of the liabilities (lending long term, refinancing short term). Second, leverage magnifies the duration of the assets. Consider a highly leveraged bank which also heavily refinances using short term instruments. Northern Rock is an example. Assume a leverage of $\lambda = 50$, $D_A = 5$ years and $D_L = 1$ year short term refinancing. Then $D_E = 250$ years follows. But this means that a moderate interest rate increase from 3 % to 3.25 % reduces the market value of the bank's equity by

$$-\frac{E'}{E}dr = \frac{D_E}{1+r}dr = \frac{250}{1.03} \times 0.25\% = 60.6\% \ .$$

How does Basel III measures act on these findings? First, the leverage ratio proves to be a strong incentive for banks to reduce leverage. Most international active banks reduce their length of the balance sheet. UBS for examples reduced its 2.8 Trillion CHF balance sheet before the crisis 2007 down to 1.4 Trillion CHF value end of 2011. Since capital - not the regulatory one - increased from 40 Billion CHF in 2008 to 54 Billion by the end of 2011, the leverage ratio of 70 in 2008 dropped to 26. Hence 1 CHF in 2008 needed to cover potential losses of 70 CHF on the asset sides while in 2011 this ratio is down to 1:26. Second, the NSFR will also impact the duration risk above as follows. NSFR is defined by

$$\text{NSFR} = \frac{\text{Available stable Funding}}{\text{Required stable Funding}} > 100\%.$$

Figure 1.18 shows the different balance sheet positions and their weight in the funding ratio.

The figure shows that on the asset side short term assets do not need stable funding since they fund themselves. The part which can not be monetarized expressed as a haircut needs stable funding. The liability side offers stable funding by the positions which are not at risk to be withdrawn. If follows that the longer the maturity of liability the higher is their value in the NSFR (they can typically be considered fully or by 100 percent in the NSFR calculation). Hence, banks will pay or favorize longer term liability business. But this reduces the difference between the duration of the asset and liability side which reduces the duration of capital by (1.4). This counter acts a similar scenario

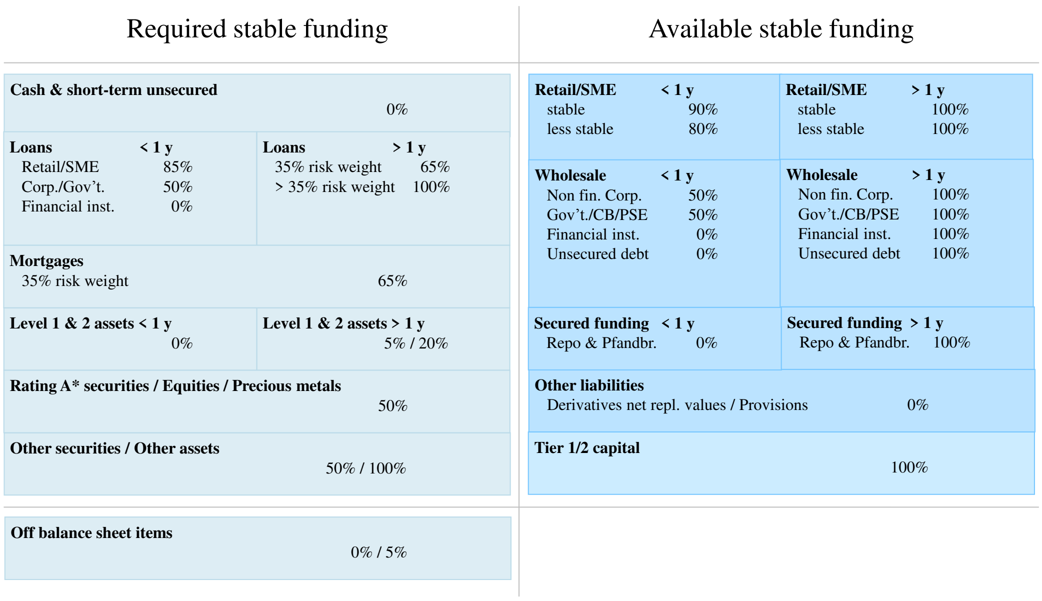


Figure 1.18: Required and available stable funding for the net stable funding ratio (NSFR). The percentage values are the factors ASF. *Source: Treasury, ZKB, 2012.*

which Northern Rock bank faced. The NSFR also reduces risk of a funding squeeze: Banks prefer to use short term funding for the longer term asset side. If a a bank comes under pressure the short term funding channel can evaporate since other banks for example in the short term markets stop providing money to the specific bank under pressure, i.e. the bank can fail to roll short term funding and faces a liquidty squeeze.

Fact 1.2.2.
- *Profitability (measured with RoE), solvency risk (measured by variance or Value-at-Risk) and funding risk are linked by the leverage factor.*

Several methods can be used to make the capital layer thicker relative to the asset layer.

- Issue new equity. This is not preferred by the actual equity holders and it is doubtful whether is works when a bank is already under stress.

- Reduce the balance sheet. Many large banks reduced considerably their balance sheet after the last financial crisis. UBS for example reduced their balance sheet from CHF 2.4 Trillion to 1.4 Trillion.

- The third possibility is to start with debt which in situation of stress for a bank is converted into equity. This defines contingent convertible bonds (CoCo bonds). CoCos solve a problem of hybrid bonds which popped up during the financial crisis. Investors considered hybrids as bonds. Regulators or rating agencies considered them due to their features to be equity pieces. This lead to conflict between the parties in particular when banks were under stress. The conflict reflects the different seniority between equity and bonds.

The new capital regulations in the Basel III require banks to deleveragea and they have a substantial impact on new types of capital instruments. **Contingent convertible equity (CoCo bonds)** have started - in particular in Switzlerand - to replace the outgoing instruments, see Oliver Wyman (2011) for details. FINMA has announced that the two big-to-fail banks UBS and Credit Suisse have to hold capital equivalent to 19 percent of RWA by the end of 2019, of which 9 percent may be held by CoCos. The subsequent issuance of USD 8 billion of CoCos by Credit Suisse was significantly over-subscribed allowing it to issue at the relatively low rate of 7.875 percent.

What is a CoCo? There are different structure: Equity conversion CoCos, where the bonds convert into equity below a certain trigger level. Write-down CoCos which involves a partial write down below a certain capital ratio are another type. There are many more variants. CoCo transacttions are also called Tier 2 Buffer Capital Notes.

We consider equity conversion CoCos. These **contingent capital** instruments have gained increasing support as a potential option to reduce the need for public bail-outs. They are bonds with principal and scheduled coupon payments that can be automatically converted into equity or written down when a predetermined trigger event occurs, enabling a fresh injection of capital into a distressed bank. Typically a trigger is a threshold level of Common Equity Tier 1 (CET1) .

If Tier 1 capital drops below a specified trigger value, the CoCo becomes equity which then increases hopefully to a level above the critical level. Using CoCos, leverage decrease in times of stress. Let $\lambda = A/E$ be the pre-crisis leverage. If conversion event is triggered, equity increases: $\lambda^* = A/(E + \text{CoCo}) < \lambda$. Swiss regulated banks (see Figure 1.17) face a new capital charge between 13.6 and 14.4 percent starting in 2012 (up from 8 percent before the crisis). Common equity Tier 1 has a 7 percent charge, i.e. about half of the total charge. If the capital buffer at 7 percent is breached, the regulator may start to impose restrictions on ordinary dividends or staff bonuses or require the institution to raise capital. CoCos are then written-down or converted to absorb losses with the aim to rapidly stabilize the bank and avoid liquidation. Cocos can be defined for different trigger values of Common Equity Tier 1 with different objectives:

- High trigger at 7 percent. The purpose of high trigger is to rapidly stabilize a bank and keep it as a going concern.

- Low trigger at 5 percent. The Coco absorbs losses but may serve mainly to facilitate resolution.

CoCos are perpetual products which pay a fixed coupon. The coupon is fixed for say 5 years and then reset every 5 years thereafter. The coupon payments are at the full discretion of the issuer of the CoCo. The issuer will be mandatorily required to defer interest payments if:

- There are not sufficient distributable items.

- A breach of a minimum capital requirement follows after a payment.

- The regulatory authority concludes this after the assessment of the financial and solvency position of the issuer.

The coupon values vary between 3 and 8 percent for recent Swiss CoCo bonds. How are the prices determined? Last year, markets were incomplete since no reference asset existed. Therefore, the common approach of relative asset pricing - i.e. using no arbitrage if there are enough assets in the market - failed. Absolute asset pricing is not used in practice. That is, one prices a CoCo by virtually defining the whole economy which matter for a CoCo and solves then the equilibrium model to obtain equilibrium prices. Innovation in such a market status is a mixture of ad hoc measure to fix a price. This includes

- to consider a reference bond which in some sense is close to the CoCo bond;

- reasoning based on experience in the structuring departments. That is there is correction of the CoCo issuer compared to the reference bond in the dimensions rating (easy), trigger event level, seniority, possible guarantees, etc.

- speaking with potential investors and trying to find out the demand function,
- try and error and comparison with similar situations.

Beside their success in some markets, some fixed income investors find CoCos in their actual form inappropriate: CoCos lack agency ratings, they violate the mandates of the many funds that are prohibited from holding equities and lack a benchmark or index against which managers can trade and measure performance.

Other actions large banks undertake to strenghten there capital basis are:

- **Mandatory convertibles**, i.e. hybrid securities (bonds linked to equities) that automatically convert to equity (stock) at a pre-determined date.
- **Exchange of deferred cash compensation** of employees into shares.
- Focus of business, i.e. selling stakes in other companies.
- Divestments and investments more liquid strategies in asset management.
- Real estate sales.
- Lower threshold deductions and additional reductions in deferred tax assets on net operating losses.

1.2.6 Credit Risk

We discuss some credit risk issues in the Basel I, II and III framework. A more detailed and updated analysis is found in Hull (2012). The goal is to obtain a feeling about the reasoning and the difficulties in the definition of regulatory standards for credit risk risk management. The first accord in 1988 defined two requirements that bank capital had to satisfy: The ratio of bank's assets to its capital had to be less than 20, i.e. there was a maximum leverage constraint which will be reintroduced under Basel III. Such a constraint already existed in many countries and was therefore already fulfilled by many banks. The second contraint is the **Cooke ratio**.. This ratio is used to calculate the risk weighted assets on- and off-balance sheet. These assets are equal to the sum of the on- and off-balance sheet risk weighted. The on-balance sheets risk weighted assets are equal to the notional amounts of the assets times a risk weighting factor w. The factor can be read-off from a table: 0 for cash, 50 percent for residential mortgage loans, etc. For the off-balance sheet positions, the **credit equivalent** is calculated for a derivative with current value E as:

$$C = \underbrace{\max(E, 0)}_{=\text{Current Exposure}} + aN$$

with a the **add-on factor** and N the notional amount. The add-on factors vary with the remaining maturity of the derivatives and the risk factor (i.e. equity, interest rates, FX, etc.). They can also read-off from a table. Note that in case of default of the counter party, $E > 0$ means an asset to the bank and therefore the bank can lose E. If E is negative, the derivative is an asset to the defaulting counter party. Hence, it reflects neither a gain nor a loss to the bank. This defines the functional form of the current exposure. Since the current exposure can grow over time, the add-on amount is introduced. This definition of the credit equivalent (i) does not properly considers how different future exposures of different derivatives can evolve, (ii) that many contracts between two counter parties could be netted and (iii) that the exposure can be reduced by collaterlization. The RWA then read:

$$\mathrm{RWA} = \sum_i w_i N_i + \sum_j w_j^* C_j$$

where the first sum is over all on-balance sheet positions and the second one over all off-balance sheet counter parties. The requirement was that the capital a bank had to keep was at least $x = 8\%$ of the RWA, where capital consisted of Tier 1 and 2 capital parts.

In the 90s of last century **netting** was introduced as follows. Netting means that the gross claim between counter parties is replaced by a single net claim. Two types of netting are of particular importance: **payoff netting** and **close-out netting**. Payment netting means that positions in the same currency and the same date are offset. This mechanism reduces settlement risk of derivative transactions. Close-out netting refers to a bilateral arrangement where both parties agree to terminate all obligations, i.e. even if they are not yet due, if default or another termination event occurs. The gross market value is added up and a single payment is obtained by the party with a negative net portfolio value. If there is at least one transaction with a positive market value then close-out netting reduces credit risk. To see this, let $E_t(i)$ be the exposure at time t of transaction i. Without any netting agreement the overall credit exposure between two parties is

$$\sum_i \max(0, E(i)) \ .$$

With close-out netting we have $\max(0, \sum_i E(i))$ and

$$\max(0, \sum_i E(i)) \leq \sum_i \max(0, E(i))$$

shows the credit risk exposure reduction due to close-out netting. This is present time view. Since netting has been successfully tested by the mid 90s the Accord of 1988 was modified to allow banks to reduce their credit equivalent totals if enforcable nettings were in place, i.e. typically if both counter parties were members of **International Swap Dealer Association (ISDA)** which signed the corresponding credit risk documents. This ment that institutions calculated the **net replacement ratio (NRR)**, i.e. the stretching factor between the exposure with and without netting:

$$\mathrm{NRR} \sum_i \max(0, E(i)) = \max(0, \sum_i E(i)) \ .$$

The credit equivalent was then changed to:

$$C = \max(\sum_i E_i, 0) + \left(\frac{4}{10} + \frac{6}{10}\text{NRR}\right)\sum_i a_i N_i \ .$$

The main change in Basel II in the quantification of the capital charge for credit risk is the new way how credit ratings of counter parties enter in the RWA calculation. Besides the Standardized Approach, which is a method based where the banks do not need to run a statistical model but where the information for the RWA calculation follows from classifying the transactions and reading off the risk weight from tables, two statistical models can be chosen: The **Foundation and Advanced Internal Rating Based Approaches (F- and I-IRB)**. We consider the A-IRB model. We do not merely state the model but provide a **microeconomic foundation** for ratings-based bank capital rules.

This section is based on Gordy (2003), Heitfield (2003) and Pykthin and Dev (2002). We consider the foundation of the capital rules in Basel II for credit risk for large institutions choosing the Advanced Internal Rating Based (A-IRB) approach, i.e. a risk sensitive model. The IRB approach relies on a bank's own assessment of its counterparties and exposures to calculate capital requirements for credit risk. The primary objectives are risk sensitivity, i.e. capital requirements under an internal model are more risk sensitive to the credit risk than in the standardized approach, and incentive compatability, i.e banks must adopt better risk management techniques to control the credit portfolio risks in order to reduce regulatory capital.

The approach is still valid under Basel III. Basel II extends the risk-based capital ratio introduced in Basel I. Three approaches to calculating risk weights: Standardized approach, IRB Foundation, IRB Advanced. Broadly, IRB Foundation considers the default risk of the debtor while IRB advanced also considers the Loss Given Default and the Exposure at Default. We focus on IRB Advanced. This approach maps - the risk-weight functions - bank-reported risk parameters to exposure risk weights Bank-reported risk parameters include, Probability of default (PD), Loss given default (LGD), Maturity (M) and Exposure at default (EAD).[29] The risk-weight functions differ by exposure classes. Classes include corporate -, industrial -, qualifying revolving - (credit cards), residential mortgages - and project finance exposures. The IRB approach is based on measures of unexpected losses (UL) and expected losses (EL). The risk-weight functions produce capital requirements for the UL portion. We identify UL with the coundary of the solvency region. Solvency is defined as above, i.e. a portfolio is solvent if the value of the assets exceeds the value of the liabilities. Risk is defined using the **Value-at-Risk (VaR)** measure stick: We set a dollar value K so that capital exceeds portfolio losses L at a one-year assessment horizon with probability α. K is the VaR for

[29] LGD is percentage of loss over the total exposure when bank's counter party goes to default, EAD is an estimation of the extent to which a bank may be exposed to a counter party in the event of, and at the time of, that counter party's default.

the given time horizon and confidence level $1-\alpha$. $K = VaR$ solves the inequality

$$P(L \leq K) \geq \alpha \ .$$

Figure 1.19 illustrates the concept.

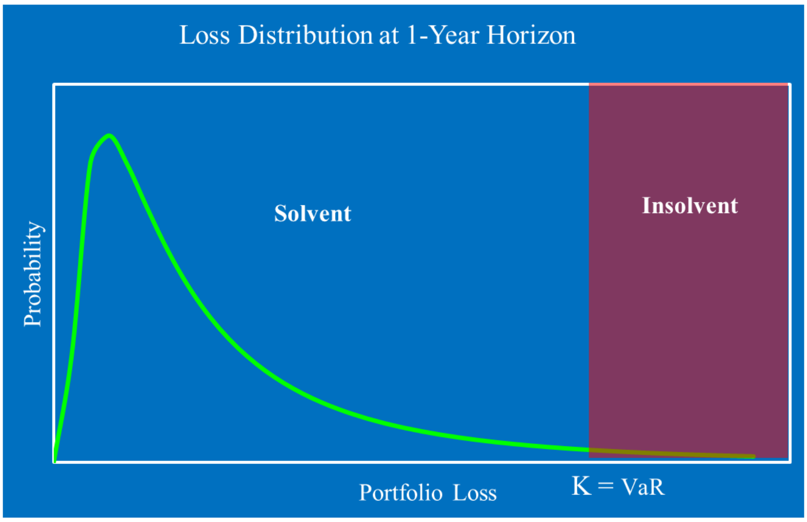


Figure 1.19: Value-at-Risk (VaR).

This shows that solvency risk measure is a limited liability concept - it is irrelevant how far away losses are from the VaR given they are larger than VaR. This concept is reasonable for an individual banking firm but it does not covers the requirements of systemic risk management which is in the scope of Basel II. In other words the framework does not takes into account the system aspect between the banks. Since the VaR directly transfers into the risk weighted assets which together with the minimum capital charge impact the necessary capital for the banks at least an add on to the VaR figures which we calculate should added reflecting the systemic risk component. Adrian and Brunnermeier (2011) propose a risk measure - CoVaR - for systematic risk. We report on this measure later. Given VaR, we require a **Decentralized Capital Rule (DCR)**. That is:

- The capital charge assigned to an exposure reflects its marginal contribution to the portfolio-wide capital requirement.
- The capital charge assigned to an exposure is independent of other exposures in the bank portfolio.
- The portfolio capital charge is the sum of charges applied to individual exposures: $K_{A+B} = K_A + K_B$.

The rationale for this rule are manyfold. First, capital charge and risk have to be aligned. Second, the capital charge attributed say to corporate banking is independent from the investment banking capital charge. This requirement avoids difficult to measure dependencies and also difficult to implement incentive schemes between different banking units. But VaR in general is not additive, i.e. the marginal contribution of a single exposure to portfolio risk depends on its correlation with all other exposures. Gordy (2003) shows that under stylized assumptions a decentralized capital rule can satisfy a VaR solvency target. The assumptions are called the asymptotic-single-risk-factor (ASRF) framework:

- Cross-exposure correlations in losses are driven by a single systematic risk factor X.
- The portfolio is infinitely-fine-grained (i.e. idiosyncratic risk is diversified away).
- Exposures loss rates $E[L|X] =: c(X)$ are increasing in the systematic risk factor X.[30]

We define the α percentile X_α of X by

$$X_\alpha = \inf\{X|P(X \leq x) \geq \alpha\}$$

and we set capital to the α percentile of L to ensure a portfolio solvency probability of α:

$$K = \inf\{k|P(L \leq k) \geq \alpha\} = \inf\{k|P(E[L|X] \leq k) \geq \alpha\} = c(X_\alpha) \ ,$$

we plug the α percentile of X into $c(X)$. The expression $\inf\{k|P(L \leq k) \geq \alpha\}$ defines the smallest loss level k such that losses exceeding k occur only with a probability of $1 - \alpha$. If we consider two subportfolios A and B it follows

$$K = c(X_\alpha) = c_A(X_\alpha) + c_B(X_\alpha) = K_A + K_B \ ,$$

i.e. capital can be assigned separately to each subportfolio. We use this ASRF capital rule to derive the IRB Advanced capital formula. That for we need a risk model for credit risk. The model used in the Accord is is **Merton's model** :[31] Obligor i defaults if its normalized asset return R_i falls below the default threshold z, i.e.

$$R_i = \epsilon_i\sqrt{1-\rho} - \rho X \leq z_i = \Phi^{-1}(PD_i)$$

with X, ϵ both standard normal, Φ the standard normal distribution function and ρ the asset correlation. X is the systematic risk factor and ϵ the idiosyncratic risk factor. The model assumes that (i) the asset return is driven by a global factor X and a obligor

[30] The assumptions read in mathematical terms:

$$P(L_A < I_A \cap L_B < I_B|X) = P(L_A < I_A|X)P(L_B < I_B|X)$$

with I_X a loss level and $X_1 > X_0 \Rightarrow c(X_1) = c(X_0)$.

[31] The bank need not employ a single model. No particular form of model is required. Analytical models are acceptable so long as they are subject to supervisory review and meet all regulatory requirements.

specific risk, (ii) the two risk factors are related by the asset correlation, (iii) the asset correlation is the same for all obligors and all type of loan products, (iv) that the threshold level defining solvency is given by the probability of default of the debtor and (v) that PD and the threshold level are related by the standard normal distribution Φ. Plugging this risk model in the conditional expected loss function for exposure i given X gives:

$$\begin{aligned} c_i(X) &= P(R_i \leq z|X)LGD_i = P(\epsilon_i\sqrt{1-\rho} - \rho X \leq \Phi^{-1}(PD_i)|X)LGD_i \qquad (1.5) \\ &= \Phi\left(\frac{\Phi^{-1}(PD_i) + X\sqrt{\rho}}{\sqrt{1-\rho}}\right) LGD_i \ . \end{aligned}$$

If we plug in the 99.9 percentile of X, which is the required percentile in Basel II, we get the core part of the Basel II capital rule using IRB Advanced:

$$K(PD, LGD) = \Phi\left(\frac{\Phi^{-1}(PD) + \Phi^{-1}(0.999)\sqrt{\rho}}{\sqrt{1-\rho}}\right) LGD \ .$$

The expected loss or capital is a function of the PD and the LGD. The crucial parameter is ρ. One assumed that his parameter measures the importance of systematic risk. After the financial crisis 2008 it became obvious that measuring systemic risk with a single parameter is not adequate. The parameter is hard wired under Basel II. The parameter was calibrated using data from various sources in Europe and the US. For corporate exposures, the parameter depends on obligor characteristics: The asset correlation declines with the obligor's PD and SMEs receive a lower asset correlation. The functional form chosen is

$$\rho(PD) = 0.12\left(1 + e^{-50\times PD}\right) \ .$$

This specification and the **maturity adjustments** are inserted in (1.6) to provide the final formula. Maturity adjustments are necessary since the capital function reflects only default losses over a one-year horizon. The parametric form of these adjustments captures that the market value of longer maturity loans are more sensitive to declines in credit quality short of default and higher PD loans are less sensitive to market value declines. The adjustments capture incremental credit risk capital due to credit migration. Figure 1.20 provides an example.

The risk weighted assets are then given by

$$\text{RWA} = K(PD, LGD) \times 12.5 \times EAD \ ,$$

i.e. with 12.5 representing 8 percent and EAD the exposure at default.

The change under Basel III which we consider here is the consideration of counter party credit risk for **derivatives**. Credit exposures on derivatives are more complicated than exposures on loans, since besides long only positions (such as for loans/bonds), positions can be short or positions can change from long to short or vice versa over time. In a long only instrument (bond), counter party risk is default risk. It can be judged by

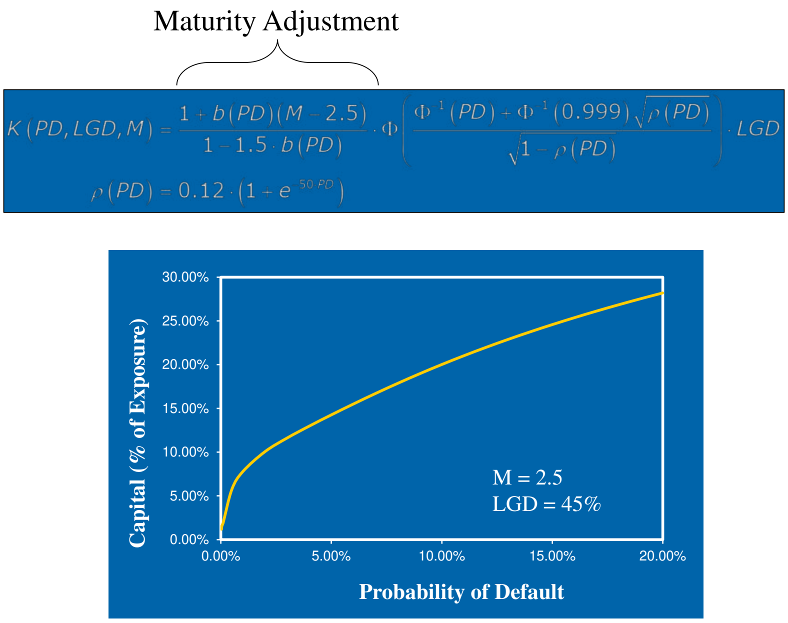


Figure 1.20: The A-IRB Capital Rule for Corporate Exposures. Maturity is $M = 2.5$ years and the LGD is 45 percent. *Source: Heitfield (2004).*

using models that can incorporate a discount curve shift. In an instrument which can have positive and negative values (swap), the instrument is either an asset or a liability. A default of a counter party then either leads to a loss or in case of a liability to an unchanged position. This makes counter party risk calculations more complicated.

The most effective ways to reduce counter party credit risk for derivatives are:

- Clearing over CCPs.
- Collateralization.
- Netting and clearing OTC.

Consider a bilateral OTC clearing arrangement between two conter parties; a trader and a company for example. The trader then calculates the **credit value adjustment (CVA)**. This is the estimate of its default if the company defaults. Since two thirds of losses in the 2008 financial crisis have been due to CVA mark to market and only about one third to actual defaults (see Nathanael (2010)), Basel III encourages institutions to include CVA mark to market future simulations in Value at Risk type measures.

'In addition to the default risk capital requirements for counter party credit risk determined based on the standardised or internal ratings-based (IRB) approaches for credit risk, a bank must add a capital charge to cover the risk of mark-to-market losses on the expected counterparty risk (such losses being known as credit value adjustments, CVA) to OTC derivatives. The CVA capital charge will be calculated in the manner set forth below depending on the bank's approved method of calculating capital charges for counterparty credit risk and specific interest rate risk.' Source: Basel III, 2011.

The CVA risk capital should account for spread and migration risks. In the first modelling approaches of the CVA and also in the bond equivalent approach of Basel III one assumes that the counter party's probability of default is independent of the dealer's exposure to the counterparty. A situation where there is a positive dependence between the two, so that the probability of default by the counterparty tends to be high (low) when the dealer's exposure to the counterparty is high (low), is referred to as '**wrong-way risk**; i.e. p and S positively depend on each other.' In case of a negative dependence one speaks about '**right-way risk**.' Consider a counter party (AIG) selling credit protection to the dealer. This situation allows for wrong-way risk as follows: When credit spreads are high, the value of the protection to the dealer is high and as a result the dealer has a large exposure to its counterparty. At the same time, the credit spreads of the counterparty are also likely to be high indicating a relatively high probability of default for the counterparty. Right-way risk tends to occur when a counterparty is buying credit protection from the dealer. There are different models to account for wrong-way risk. The α multiplier approach is the simplest one. Another method used is to set $S(t)$ equal to the present value of the exposure that is k standard deviations above the average exposure for some k.

We consider **interest rate swaps (IRS)** as an example for CVA calculation; we follow Stein (2012). Assume a 3 year IRS in a flat interest rate environment. With $V(t)$ the value of the risk free swap at time t, R the constant recovery value and τ the default time of counter party B. If B defaults at time τ, the payoff for A is $RV(\tau)$ if $V(\tau) > 0$ and $V(\tau)$ if $V(\tau) < 0$. Summarizing, the payoff for A is

$$R \max(V(\tau), 0) + \min(V(\tau), 0) = V(\tau) - (1 - R) \max(V(\tau), 0).$$

The second term is an optionality, i.e. to enter into a swap contract at default time which pays what is then left $V(\tau)$ or 0. Such an option an a swap is called a **swaption**. The decomposition shows that the exposure is a sum of a risk free part and a risky one. What are the costs of the loss $(1 - R) \max(V(\tau), 0)$? Following first pricing principles the loss - **CVA** - is equal to the risk neutral expected value under a numeraire measure N, i.e.:

$$\text{CVA}(0) = N(0)(1 - R)E[\frac{\max(V(\tau), 0)}{N(\tau)}\chi_{\tau<T}] \ .$$

Using the delta function[32], we can write[33]

$$\text{CVA}(0) = N(0)(1 - R) \int_0^T E[\frac{\max(V(s), 0)}{N(s)}\delta(s - \tau)]ds \ .$$

This shows that the CVA is a product of call $\frac{\max(V(s),0)}{N(s)}$ and the default event $\delta(s - \tau)$. If they are **independent**, the expectation factors. The CVA then becomes the product of the current value of the swaption $S(s)$ to enter into the remainder of the swap at time s and the default time probability density function $p(s)$, i.e.

$$\text{CVA}(0) = (1 - R) \int_0^T S(s)p(s)ds \ , \ \text{CVA}(0) = (1 - R) \sum_{j=0}^{n} S(j)p(j)$$

with p_j the probability of a default or **credit spread** unwind between times t_{j-1} and t_j and $S(j)$ is evaluated at the midpoint of the interval $[t_{j-1}, t_j]$. There are many subtleties in the calculation of the CVA integral. We refer to the literature for details, see Stein (2012) and references therein. The analysis shows that CVA are themselves derivatives and must be managed similarly to other derivatives, they are complex derivatives since it is contingent on the **net value of the portfolio** of derivatives outstanding with that counterparty. Calculating CVAs is computationally intensive. Reuters (2008) reported

[32]The delta function is not a function in a strict mathematical sense but a so-called generalized function. Nevertheless, physicists use since decades to work intuitiively with this and other generalized functions. In economics, the delta function describes an Arrow-Debreu security in continunuous time: $\delta(x - y)$ is zero if $x \neq y$ and 'infinite' if $x = y$. One uses that $\int f(x)\delta(x - y)dx = f(y)$. Setting f equal to the identity, the delta function gives all its weight to a single point.

[33]The expectation is under a risk neutral measure to avoid arbitrage.

that, at the time of its failure Lehman, had about 1.5 million derivatives transactions outstanding with 8,000 different counterparties. It had to calculate 8,000 different CVAs each with an average number of 200 derivative transactions.

We consider the Basel III formula for the CVA. An approximation of the average hazard rate between time 0 and time t is $s(t)/(1-R)$ with $s(t)$ the credit spread of the counter party at tenor t. The an estimate of the probability of no default between times 0 and t_i is $\exp(-s_i t_i/(1-R))$. This implies for the probability of default

$$\tilde{q}_i = \exp(-s_{i-1}t_{i-1}/(1-R)) - \exp(-s_i t_i/(1-R)) \ .$$

This is one part of the Basel III formula for the CVA calculation in the advanced measurement approach. The second part is to replace $1-R$ by observed market LGDs, i.e. the loss given default of the counterparty and should be based on the spread of a market instrument of the counterparty. Finally, the swaption exposure $S(j)$ is represented as the **average expected expsoure** at the dates $j-1$ and j of the swaption: $S(j) \rightsquigarrow \frac{1}{2}8EE(i-1)D(i-1) + EE(i)D(i))$ with D the discount factor and $EE(i)$ the expected exposure to the counterparty at revaluation time i. The use of the above CVA formula requires that the bank uses an IRB model and an internal market risk model for interest rate risk. What happens to all other smaller banks? As for default risk, there is a **standardized CVA** formula. The capial charge reads

$$\begin{aligned} K &= 2.33\sqrt{h}\sqrt{A+B} && (1.6)\\ A &= \left(\sum_i \frac{1}{2} w_i (M_i EAD_i^{tot} - M_i^{hed} B_i)^2 - \sum_{k \in ind} w_k M_k B_k\right)^2 \\ B &= \sum_i 0.75 w_i^2 * (M_i EAD^{tot} - M_i^{hed} B_i)^2 \end{aligned}$$

with: h is the one-year risk horizon, w_is the weight applicable to counterparty i, EAD_i^{tot} the total exposure at default of counter party i (after netting and including collateral), B_i is the notional of purchased single name CDS hedges referencing counterparty i and used to hedge CVA risk, ind refers to indices of indizes (CDS, counter party weights), M_i is the effective maturity of the transactions with counterparty i and M_i^{hed} is the maturity of the hedge instrument with notional B_i. We refer to Pykhtin (2012) for a foundation of the formula (1.6).

1.2.7 Leverage

We have encountered the importance of leverage for RoE and funding. But leverage is a broader concept in finance and accounting. In particular we want to tackle some aspects of leverage and asset value which were important in the 2008 financial crisis. We consider four different situations:

- Leverage for different types of investors.

- Target leveraging and asset fluctuation
- The impact of exogenous shocks on individual balance sheets and its interplay with leverage and regulatory capital charge requirements.
- Leverage and risk measurement from a systemic risk point of view.

1.2.7.1 Leverage and Investment

We consider the role of **leverage for different types of investors**. We have

- an optimistic or aggressive investor;
- a pessimistic investor.
- a long-only investor such as a pension fund.

We set S for the price of the risky asset (stock), ϕ the number of stocks in the portfolio, $A = \phi S$ the value of the risky asset, E capital and C cash. Figure 1.21 shows the balance sheets at a fixed date.

Types of Investors

Aggressive	Assets	Liabilities	
Stocks	A	-C	Debt
		E	Equity

Pessimistic	Assets	Liabilities	
Cash	C	A (<0)	Debt
		E	Equity

Long only	Assets	Liabilities	
Cash	C		
Stocks	A	E	Equity

Homeowner

Time 0	Assets	Liabilities	
House	555.5	500	Mortgage
Cash	144.5	200	Equity
Total	700	700	

Time 5y	Assets	Liabilities	
House	388.5	350	Mortgage
Cash	0	5.5	Debt
		33	Equity
Total	388.5	388.5	

Figure 1.21: Left Panel: Balance sheet for the three types of investors.*Source: H. Shin, 2011.* Right Panel: Balance sheets for a home owner.

The aggressive investor borrows cash $C < 0$ to buy the risky asset. The pessimistic one is short the risky asset ($\phi < 0$). He makes a profit if the risky asset drops. For the aggressive investor the asset side of the balance sheet is at risk; for the pessimistic one the liability side. The third type holds both the risky asset and cash in his portfolio. The leverage of the three types is different:

- The optimistic investor has a leverage $\lambda := A/E > 1$. The more aggressive he is, the higher borrowing, the more risky assets and the higher leverage.
- The leverage of the pessimistic investor $\frac{C}{E} = \frac{E-A}{E} = 1 - \lambda$ is smaller than 1. Note that $A < 0$.
- The long only investor is not leveraged. The ratio between assets and capital is always $\frac{C+A}{E} = 1$.

1.2.7.2 Target Leveraging and Asset Fluctuation

We discussed the connection between leverage, RoE and risk in Section 1.2. Consider a **single home owner** which bought a house for a price of 555.5 in a currency. Financing was made with a fixed rate 5y mortgage with 90 percent LTV (loan-to-value) i.e. of 500, see Figure 1.21. The balance sheet after the house was bought shows that the home owner has a capital of 200 which leads to the residual cash account of 144.5. Assume that house prices drop by 30 percent in 5 years. The value of the house is then 388.5. The maximum LTV ratio of 90 percent implies that the value of the mortgage is bound to $0.9 \times 388.5 = 350$. Therefore after 5y the house keeper has to inject capital of $500 - 350 = 150$. This is a large amount. In particular, the cash reserves vanishes and the home owner faces additional debt of 5.5. In terms of leverage $\lambda = A/E$, he starts at initiation with a leverage of $3.5 = 700/200$. After the price decay leverage is $11.8 = 388.5/33$. This is not an example for subprime mortgages since this home owner hardly face a equity/house price ratio of 200/555.5. Such an increase in leverage triggers home owners default since the probability of insolvency $P(A < L)$ increase if leverage λ increases. The impact of changing interest rates on leverage is much weaker as it is for changing house prices. To understand the difference of falling house prices or raising interest rates on leverage we proceed as follows. Setting $xA = A'$ with $x < 1$ the new asset value after a drop of house prices where the factor x also scales capital the new leverage after the shock λ' reads up to first order (Taylor expanding the denominator (capital) up to first order):

$$\lambda' = \lambda + \lambda x \frac{D}{E} + \text{higher orders} .$$

To consider the impact of an increase of the interest rate we by yD with $y > 1$ the increase in debt due to higher interest rate payments. Using this in the definition of the leverage, i.e. reducing capital in the denominator and again expanding the denominator up to first order one gets:

$$\lambda' = \lambda + \lambda y + \text{higher orders} .$$

Assuming that an investor has low equity D/E is close to A/E. Hence in the case of a house price shock

$$\lambda' \sim \lambda + \lambda^2 x ,$$

i.e. a quadratic growth. Using figures $\lambda = 4$ and house price decrease of 20 percent ($x = 0.8$) versus an increase of interest by 3 percent, $y = 1.03$, leads to a new leverage of approximatively 16.8 for the first case and of 9 in the second case.

What happens to leverage as **total assets fluctuate**? Since leverage is inversely related to total assets ($\lambda = \frac{A}{A-D}$), when the price of the house goes up, the net worth increases, and the leverage goes down. The negative relationship between total assets for households and leverage is clearly borne out. But this observation does not holds true for companies, commercial banks and brokers or investment banks, see Figure 1.22.

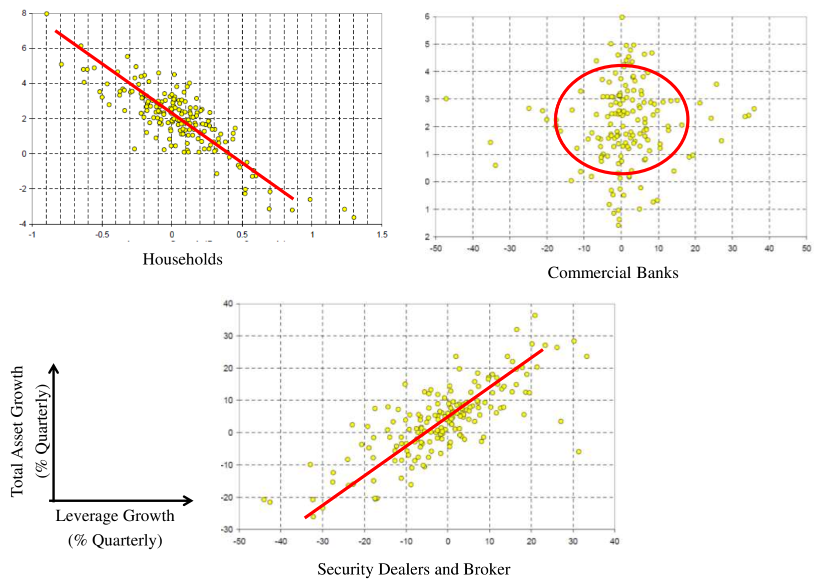


Figure 1.22: Relationship between assets and leverage. *Source: Adrian and Shin, 2007.*

For commercial banks a large number of the observations line up along the vertical line that passes through zero change in leverage, i.e. commercial banks are targeting a fixed leverage ratio. Financial institutions manage their balance sheets actively for several reasons. They attempt to manage the key financial ratios so as to hit credit rating targets and the cost of capital. Their models of risk and economic capital also dictate active management of their balance sheets: Economic capital is related to performance measures such as RoE. RoE increases with an increasing leverage. Increasing RoE can lead in banks using this performance measure to increasing compensation. The scatter chart for security dealers, brokers and investment banks show the reverse pattern of that for households. There is a strongly **positive** relationship between changes in total assets and changes in leverage. In this sense,

Fact 1.2.3. *Leverage is pro-cyclical.*

To illustrate this let $\lambda = A/E$ be the leverage before a change in asset values. After a change of $A \to A + \delta A$ to restore the leverage, debt has to be changed to $D \to D + \delta D$ to trade on the asset side. I.e.

$$\lambda' = \frac{A + \delta A + \delta D}{E + \delta A} = \lambda$$

implies

$$\delta D = \delta A(\lambda - 1) \ .$$

If asset prices fall ($\delta A < 0$), securities are sold. Contrary, an increase in the asset prices leads to positive demand to buy assets. Hence, to target leverage leads to upward-sloping demands and downward-sloping supplies. This atypical pattern of demand and supply is reinforced if leverage λ is high. But leverage is high if asset prices are high i.e. in a boom period and vice versa for a bust period, see Figure 1.23.

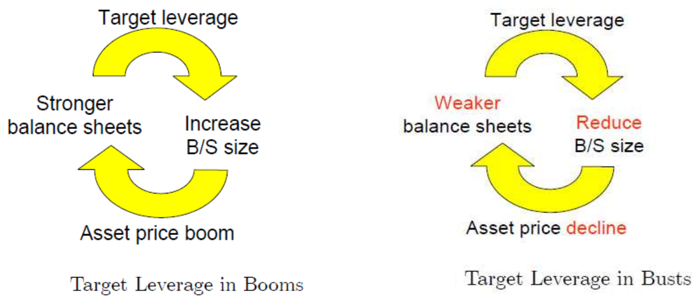


Figure 1.23: Target leverage in booms and busts.. *Source: Adrian and Shin, 2007.*

1.2.7.3 Leverage, Shocks and Regulatory Capital Charge

We consider the impact of **exogenous shocks** on the balance sheet, the regulatory charges and its interplay with leverage. We follow Mele (2011). Consider again a balance sheet where today we have:

- Asset side: Asset value A plus cash C.
- Liability side: Debt D plus capital E.

We assume that equity has to be larger than a given percentage of the asset value, i.e. $E \geq xA$ with x a percentage value. In Basel II $x = 8\%$ and A corresponds to the risk-weighted-assets. For simplicity we assume that equality $E = xA$, i.e. the bank just satisfies the minimal capital requirements (which is reality would not be accepted by the regulators, i.e. they would intervene before this happens). Therefore the percentage value x equals inverse leverage $1/\lambda$. An increase in the capital ratio x is equivalent to a decrease of the leverage ratio. Assume a shock $-\Delta A$ on the asset value where the asset value A and capital E are both reduced by this amount. Then,

$$E - \Delta A = x(A - \Delta A)$$

has to hold to maintain the regulatory requirement. This is impossible since equity falls by a larger percentage than the asset side due to leverage, i.e. $\frac{-\Delta A}{E} < \frac{-\Delta A}{A}$ since $E < A$. Two solutions are available to the financial institution to restore regulatory capital requirements: (i) to inject fresh capital; (ii) to sell some of the risky assets. The first solution is not quite viable in the short-run. Let us analyze the second solution. We are looking for a quantity X of the risky asset to sell, such that after a reduction in the asset value the capital ratio target is still met. The balance sheet reads:

- Asset side: Assets $A - \Delta A - X$, cash $C + X$
- Liability side: Debt D plus capital $E - \Delta A$.

To maintain the regulatory capital adequacy rule, X solves

$$E - \Delta A = x(A - \Delta A - X) \ .$$

Solving, the number of risky assets to sell is proportional to the percentage loss in their value:

$$X = \Delta A \left(\frac{1}{x} - 1 \right) = \Delta A \left(\lambda - 1 \right) \ .$$

If we use this value for X, leverage before and after the shock are the same, i.e.

$$\lambda' = \frac{A - \Delta A - X}{E - \Delta A} = \frac{A}{E} = \lambda \ .$$

There is relation between leverage and risk measurement, i.e. the Value-at-Risk (VaR) at a confidence level α.[34] We assume that equity equals the VaR, i.e. the bank has no reserve capital but all of its capital is allocated to measurable risks. Then

$$E = \text{VaR} = v \times A$$

[34] VaR is the smallest non-negative number such that

$$P(A - A_0 < -\text{VaR}) \leq 1 - \alpha$$

with A_0 the base asset level and α the confidence level.

with v the value-at-risk per dollar of assets implies that

$$\lambda = \frac{1}{v} = \frac{1}{x} \ .$$

Hence in this simple setup the regulatory capital asset ratio equals the value-at-risk of one dollar which is inversely related to the leverage.

Consider an asset price shock $A \to A + \delta A$, which also shocks the value-at-risk $v \to v + \Delta v$. Then

$$\frac{1}{\lambda} \sim v(1 - \frac{\delta A}{A}) + \Delta v(1 - \frac{\delta A}{A}) + \frac{\delta A}{A} > v$$

holds if $E = vA$ and if we develop the quotient in a power series dropping terms of second order such as $(\delta A)^2/A^2$. It is an empirical regularity that if assets drop, volatility and hence VaR increase (negative asset leverage). Hence, $\delta A < 0$ and $\Delta v > 0$ hold. This implies $\frac{1}{\lambda} > v$. Leverage is decreasing using a value-at-risk measure if markets fall.

The above models do not consider that there can be a **price impact** if say fire sales of assets are needed to restore regulatory capital requirements. Cifuentes, Ferruci and Shin (2005) consider the possible impact of a shock on asset prices (spiralling down). They explore liquidity risk in **a system of interconnected financial institutions** when these institutions are subject to regulatory solvency constraints and mark their assets to market. The institutions are interconnected via their balance sheets - a bank's i asset are a bank j's liabilities. There are n banks which form the interbanking market. The liability of bank i to bank j is denoted D_{ij} with $\bar{D}_i = \sum_j D_{ij}$ the total liability face value of bank i. This notional value can be different from the market value D_i of bank i's interbank liabilities. We assume equal seniority of all claims. Then bank's payments are proportional to the notional liability, i.e. $\bar{D}_i \pi_{ij} = D_{ij}$. $\sum_j \pi_{ji} D_j$ are all payments bank i receives from all other banks, i.e. it is an asset value. Furthermore, there is an illiquid asset A_i^{ill} with price p.[35] The equity value for bank i is then equal to the difference between the two asset values and the liability value:

$$E_i = pA_i^{ill} + \sum_j \pi_{ji} D_j - D_i \geq 0 \ .$$

The positivity follows from limited liability of the bank. Priority of debt over equity implies that equity value is strictly positive only when bank i's payment is equal to its notional obligation, i.e. $\bar{D}_i = D_i$. Therefore bank i payment satisfies

$$D_i = \min\{\bar{D}_i, pA_i^{ill} + \sum_j \pi_{ji} D_j\} =: \min\{\bar{D}_i, w(p) + \sum_j \pi_{ji} D_j\}$$

where $w(p)$ is the marked-to-market value of the illiquid asset. From a secured banking view,

[35] The authors add a liquid asset too. Since this is not necessary for our discussion we leave this asset.

- D_i is the debt capacity;
- $w(p) + \sum_j \pi_{ji} D_j$ is the collateral value;
- E_i is the haircut.

Writing Π for the payoff matrix of the interlinked claims, the vector D of all clearing payments satisfies the system of n equations:

$$D = \min\{\bar{D}, w(p) + \Pi' D\} \ .$$

This is a system of **realized debt values**: Symbols which are not notional values are realized values at future date. A vector D which satisfies this equation is a fixed point of the function $f(D) = \min(\bar{D}, w(p) + \Pi' d) = D$. By Tarski's fixed point theorem, there is at least one fixed point or **clearing vector** D. A sufficient condition for the existence of a unique fixed point is that, first, the system is connected in the sense that the banks cannot be partitioned into two or more unconnected sub-systems, and that there is at least one bank that has positive equity value in the system, see Eisenberg and Noe (2001). We assume that uniqueness holds: There exists a unique clearing vector $L(p)$ as a function of the illiquid asset price p.

We add to the system the **supervisory capital adequacy ratio**, i.e. a lower bound on the capital asset ratio of the bank. The constraint is given by

$$\frac{pA_i^{ill} + \sum_j D_j \pi_{ji} - D_i}{pA_i^{ill} - ps_i + \sum_j D_j \pi_{ji}} \geq x$$

with x the pre-specified capital ratio, s_i the units of the illiquid asset sold by bank i. The numerator is the equity value of the bank calculated in terms of the expected payments. The denominator is the marked-to-market value of its assets after the sale of s_i units of the illiquid asset. Credit risk is not considered. The assumption of the model is that assets are sold for cash, and that cash does not attract a capital requirement. By selling its assets for cash, the bank reduces the size of its balance sheet and thus reduces the denominator, i.e. it increases the capital ratio. We assume that the bank cannot short sell the assets, i.e. $s_i \in [0, A_i^{ill}]$.

Definition 1.2.4. *An* ***equilibrium*** *is a triple (D, s, p) with a clearing vector D, a vector of sales of illiquid asset s and the price p of the illiquid asset such that:*

- *for all banks i, the smallest sale that ensures that the capital adequacy condition is satisfied s_i the limited liability of equity holders, and the priority and equal seniority of the debt holders holds, i.e. D satisfies the clearing fixed point equation..*
- *For all banks i, s_i is the smallest sale that ensures that the capital adequacy condition is satisfied. If there is no value of $s_i \in [0, A_i^{ill}]$ for which the capital adequacy condition is satisfied, then $s_i = A_i^{ill}$: Either the bank is liquidated altogether, or its sales of illiquid assets reduce its assets sufficiently to comply with the capital adequacy ratio.*

- *There is a downward sloping inverse demand function $d^{-1}(\cdot)$ such that $d(p) = s(p)$, i.e the price of the illiquid asset is determined by the intersection of a downward sloping demand curve and the vertical supply curve given by aggregate sales $s(p) = \sum_i s_i(p)$.*

By re-arranging the capital adequacy condition, we get

$$s_i(p) = \min\left(A_i^{ill}, \frac{D_i - (1-x)(\sum_j D_j \pi_{ji} + pA_i^{ill})}{xp}\right) .$$

Since interbank payments D_{ij} are all functions of p, s_i itself is a function of p. Since each $s_i(\cdot)$ is decreasing in p, the aggregate sale function $s(p)$ is decreasing in p. Assuming the inverse demand curve for the illiquid asset

$$p = e^{-\alpha \sum_i s_i} \ , \ \alpha > 0$$

the maximum price $p = 1$ follows for zero sales. This is the status quo price where the banking system has not suffered any adverse shock. We assume that if the entire endowment of illiquid assets in the system is sold, there is at least one bank that has positive equity value, i.e. for $\hat{p}$ the price of the illiquid asset when the entire endowment of the illiquid asset is sold supply at this price level is dominated by the demand. Figure 1.24 shows the price adjustment process starting with an intimal price $p = 1$ and a shock which leads to a price p_0 where forced sales of the banks puts quantity $s(p_0)$ on the market. This pushes the price further down to $p_1 = d^{-1}(s(p_0))$ which leads to further sales and a supply $s(p_1)$. This process then continuous until the intersection point between supply and demand. This shows how a shock can lead to a **downward spiraling** of the illiquid asset prices in the interbanking market.

To consider the leverage and risk measurement in this setup we write for the marked-to-market value of assets of bank i (see Figure 1.24)

$$A_i = A_i^{end} + \sum_j \pi_{ji} D_j$$

with A_i^{end} the loans to end users. The balance sheet is $A_i = E_i + D_i$. Using matrix notation for the n banks in the system, the vector of debt values $D \in \mathbb{R}^n$ satisfies

$$D = \Pi D + A - E \in \mathbb{R}^n$$

with Π the matrix with entries π_{ij}. Defining leverage $\lambda_i = A_i / E_i$ for bank i and Λ the diagonal matrix of all leverage factors, we get

$$A = E + E(\Lambda - \mathbb{I})(\mathbb{I} - \Pi) \tag{1.7}$$

with $\mathbb{I}$ the identity matrix. This follows from $D_i/E_i = \lambda_i - 1, D = E(\Lambda - \mathbb{I})$.

Fact 1.2.5. *Total lending to the end-user borrowers depends on the interaction of the distribution of equity E in the banking system, the profile of leverage Λ and the structure of the financial system Π. Total lending to end users is increasing in equity and in leverage.*

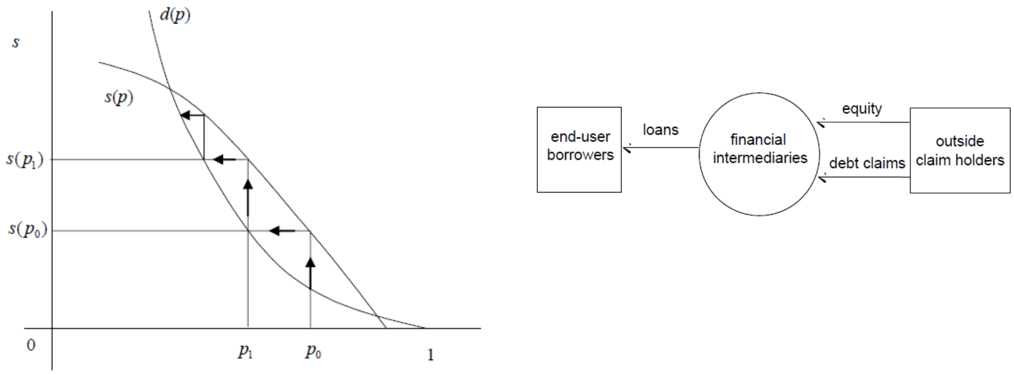


Figure 1.24: Left Panel: The price adjustment process can be depicted as a step adjustment process in the arc below the $s(p)$ curve, but above the $d(p)$ curve. Right Panel: Stylized financial system. *Source: Cifuentes et al. 2005.*

To understand the impact of the structure of the banking system Π on lending we write $z = (\mathbb{I} - \Pi)u$ with $u' = (1, 1, \ldots, 1)$. Therefore $z_i = 1 - \sum_j \pi_{ij}$ is the proportion of bank i's debt held by the outside claimholders. Multiplying (1.7) by u total lending to end users reads:

$$\sum_i A_i = \sum_i E_i + \sum_i E_i z_i(\lambda_i - 1) \ . \tag{1.8}$$

This is the **balance sheet identity for the financial sector** where all the claims and obligations between banks have been netted out. The first term on the right-hand side is total equity of the banking system. The second one is total funding to the banking sector provided by the outside claim holders. Hence, credit supply to end-users must come either from the equity of the banking system or the funding provided by non-banks.

We consider the **financial system leverage**. We construct a financial system where the aggregate equity, lending and leverage are all unchanged but where the debt to equity ratio of all individual banks is μ times as large, i.e. $D' = \mu D, E' = E, \Pi'$ where the sum over the entry of row j are $1 - z_j/\mu$. Assume first that all values are **face values**. The balance sheet identity in this new system reads

$$A' = E' + L'(\mathbb{I} - \Pi') \ .$$

Multiplying with the vector u leads for aggregate lending

$$\sum_i A'_i = \sum_i A_i \ .$$

The aggregate leverage is **unchanged** but the debt to equity ratio of all individual banks is larger in the second financial system. There is only one restriction $1 - z_j/\mu > 0$, i.e. a lower bound on μ. But there is no upper bound. This proves that there exist a financial system where aggregate leverage is unchanged but individual leverage can be made arbitrary large. The intuition is that a high individual leverage remains within the banking sector but **does not swaps to the ultimate creditor sector**. The analysis is unchanged when using **market values** with one exception: Since the market value of debt cannot be larger than the market value of the assets, there is an upper and a lower bound for μ. Leverage of the systems λ is given by:

$$\lambda = \frac{\sum_j A_j}{\sum_j E_j} = 1 + \frac{\sum_j E_j z_j(\lambda_j - 1)}{\sum_j E_j} \ .$$

Fact 1.2.6. *Total system leverage increases if the amount from outside financing z increases.*

This approach can be enriched by introducing risk management issues, i.e. value-at-risk. Shin (2006) provides this extension.

1.2.8 Regulation of OTC Derivatives

OTC derivatives are bilateral agreements to hedge risks or bet on prices. They are taylor-made and there was little regulation in the past. This lead also to lack of transparency about the counter parties. Since these markets grew heavily in the near past the lack of information led to uncertainty about the risks of these markets for the counter parties but also for the whole financial system, see Figure 1.25 for the size of the markets. If one assumes a notional amount of about 600 trillion USD in the OTC markets and if one compares this figure with the world wide GDP of about 55 trillion USD one might wonder about the rationale why for each dollar GDP there exist 11 dollars of protection transactions.

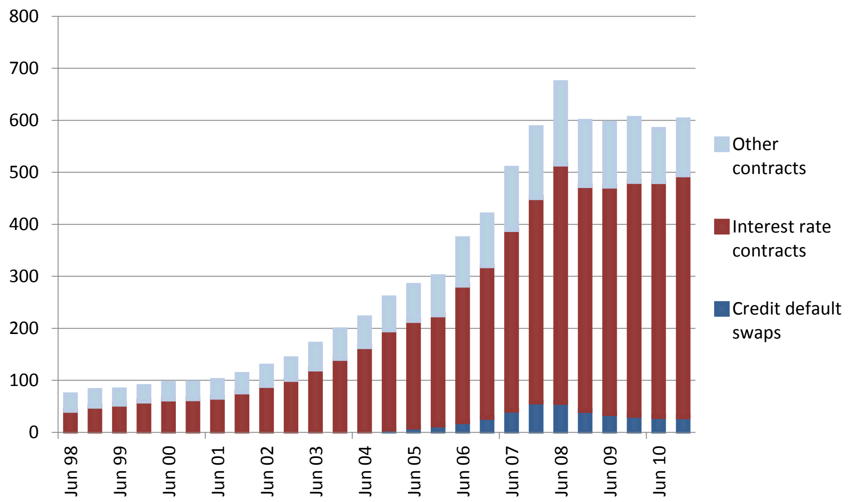


Figure 1.25: The size of the OTC derivatives market in terms of notional amount outstanding (USD trillion).*Source: BIS (2011)*

The size in these markets is a poor risk figure since several risk mitigating methods exist. The most prominent are the Master Agreement and Credit Support Annex (ISDA). These agreements reduce counter party risk in the bilateral trades. Major instruments in doing these are netting agreements and collateralization. Although this is effective on a bilateral level due to the privacy of the trades (i) it is not transparent from a system point of view and (ii) it does not provides a global view on the whole network of interlinked counter parties. Due to this lack of reporting and supervision a problem of undercollateralization can occur. A prominent example was AIG which served as protection seller for many investment banks, i.e. AIG obtained a premium from the banks in exchange of credit protection by AIG if the reference entities default. AIG faced a liquidity crisis due to a downgrade of its rating. Such a downgrade forced AIG to post additional collateral with its trading counter parties. Furthermore the credit protection sold lost in value since the market for CDOs (Collateralized Debt Obligations) were quasi evaporating. This led to a first bailout by the FED. There were others to follow with a total amount of 182 Billion USD. The first of bailout is an example of risk structuring and risk transfer. The FED created an 85 billion credit facility in exchange for the issuance of a stock warrant to the Federal Reserve Bank for 79.9% of the equity of AIG. Hence, the FED exchanged credit and market risk. It took the credit risk of AIG in exchange for a future participation in the firm value.

A main requirement to reduce some the mentioned risks in OTC transactions is to **clear** the contracts via a **central counter party (CCP)**. The idea is that the contracts are no longer between say two investment banks but between each investment bank and a clearing house. This has the following consequences:[36]

- Standardized OTC derivatives (interest and credit derivatives, the clearing of FX contracts is an open issue) are traded on exchanges or electronic platforms.

- The contracts are cleared, i.e. the clearing house applies the margin process. Figure 1.26 shows the logic of client clearing.

  This means that on a daily basis difference in value of the contracts are settled. Using such a procedure it is not possible that positive or negative value accumulate over time. From a risk perspective OTC trades are only partially secured. In the future trades will be overcollateralized and all clearing members are liable to a certain extend if clearing members default. More precisely, the 'risk waterfall' method applies. A CCP has a multi-layer capital structure to protect itself and its members from losses due to member defaults. The following types of collateral will be held (we follow Arnsdorf (2011), Pirrong (2009) and ZKB (2012) in the sequel):

  – Variation Margin: Variation margin is charged or credited daily to clearing member accounts to cover any portfolio mark-to-market (MtM) changes.

  – Initial Margin: Initial margin is posted by clearing members to the CCP. This is to cover any losses incurred in the unwinding of a defaulting member's portfolio. Typically the margin is set to cover all losses up to a pre-defined confidence level in normal market conditions.

  – CCP Equity: A CCP will have an equity buffer provided by shareholders.

[36]The consequences meet the statement of the G20 in their Pittsburgh Summit 2009: *'All standardized OTC derivative contracts should be traded on exchanges or electronic trading platforms, where appropriate, and cleared through central counter parties by end-2012 at the latest. OTC derivative contracts should be reported to trade repositories. Non-centrally cleared contracts should be subject to higher capital requirements.'*

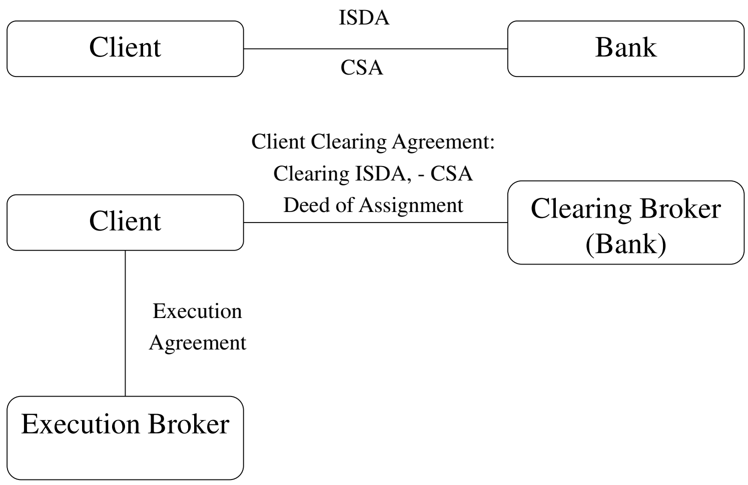


Figure 1.26: Client clearing under the existing ISDA and CSA setup and under the new clearing requirements.

  - Default Fund (funded): Every member contributes to the clearing house default fund. This acts as a form of collective insurance for uncollateralised losses.
  - Default Fund (unfunded): Each clearing member is usually committed to providing further funds if necessary. The maximum amount of additional funds that can be called upon depends on the CCP. In some cases the liability is uncapped.

- Clearing the OTC contracts via a clearing house changes the bilateral random network topology of all contracts to a star shaped one.

Participants in the OTC markets have three options: They can do nothing which means that they will stop to use say interest rate swaps. They can evaluate a clearing broker. This requires to invest into updating collateral risk management, accounting workflows and the redefinition of backoffice post trade workflows. The third alternative is to become a member of a clearing house. This allows for large flexibility to offer client clearing. Besides investment costs to change workflows one has also to consider the amounts payable into the default fund. To understand the alternatives, we consider them in an example regarding capital requirements. This example is an extension of an example developed by ZKB, 2012. The regulatory capital charge for OTC derivatives increases since bilateral OTC derivative contracts have to satisfy the Credit Value Adjustment (CVA) requirements and the risk weights of cleared OTC derivatives increase from 0% to 2% . Cleared transactions will receive lower risk weights and can abstain from CVA if portability is given with a high probability. It is not yet defined what 'high probability' means. We consider the following portfolio of CHF-IRS.

Notional	TtM	Add On	NPV	EAD
500	9m	0%	5.03	5.03
700	4y	0.5%	-20	-16.5
400	10y	1.5%	40	46

The currency figures are in million CHF. The EAD is calculated using the standardized approach, i.e.

$$\text{EAD} = \text{Notional x Add On} + \text{NPV} .$$

The portfolio has a credit equivalent after netting is 32 Mio. CHF. With a rating of AAA to AA- this gives the 8% capital charge of 512'000 CHF. A lower rating of A+ to BBB- gives in the standardized approach the amount of 1.28 Mio.CHF. The nominal weighted TtM is 4.5 years. For a rating of AAA to AA the CVA capital is 2.3 Mio. CHF and 2.7 Mio. for a A rating.

If one clears the portfolio collaterization applies. The clearing client pays an initial margin of 4 Mio. CHF and receives a variation margin of 25.03 Mio. CHF. The capital requirements ...

- using two clearing brokers is 17'500 CHF by applying the risk weight of 2 percent.
- using a single AAA clearing broker is 175'000 CHF by applying the risk weight of 20 percent and the minimal CVA capital requirements are 510'000.
- using a single A clearing broker is 440'000 CHF by applying a risk weight of 50 percent and the minimal capital requirements for the CVA are 582'000 CHF.

This calculation shows the trade off between capital costs and costs for a second clearing broker. CVA should be part of the trade valuation but calculated separately because of portfolio effects. If one uses cleared OTC the risk weights increase from zero to 2 percent.

Pirrong (2009) shows the following effects of CCPs:

- In conditions of complete information, a CCP can improve welfare by allocating default losses more efficiently than in a bilateral network. A CCP can reduce the frequency and severity of default losses that hedgers suffer in these high marginal utility states, thereby improving welfare. The formation of a CCP also affects equilibrium prices, quantities, and profits.
- *'A CCP affects the distribution of default losses among market participants. Netting effectively gives derivatives counter parting a priority claim on assets of an insolvent counter party, and therefore transfers wealth from other creditors to derivatives counter parties. Moreover, CCPs insure non-members against losses arising from a dealer default, thereby effectively transferring the burden of these losses from non-members to the financial institutions that are members of CCPs.'*
- *'Due to the distributive effects of clearing, and its effect on the pricing of default risk, it is not necessarily true that formation of a CCP reduces systemic risk. Indeed, it can increase systemic risk under some circumstances.'*

These results are all based on an economic model. Other approaches such as Duffie and Zhu (2011) show that there is ambiguity to which extend CCPs reduce systemic risk.

OTC regulation applies to banks, insurance companies, asset managers and corpoartes with large OTC sizes. Pension funds are so far not included. Clearing houses are London Clearing House for interest rate swaps, ICE or CME for credit default swaps. The platforms where OTC trades are executed are so called multibank platforms.

Although bilateral counter party risk is reduced and overall transparency is gained with this initiative some issues are still open. First, the yet scarce resource of collateral will become more expensive due to the overcollateralization. This will make some OTC contracts less attractive.

1.3 Technology and Technological Shocks

Shocks to technology provide a supply-side explanation for the timing of some innovations. Advances in information technology support sophisticated pooling schemes which we observe in securitization. IT and improvements in telecommunications (internet) has facilitated a number of innovations, including new methods of underwriting securities, new methods of assembling portfolios of stocks, new markets for securities and new means of executing security transactions. IT innovations - storage capacity, performance - often allow intellectual finance innovation to become operational. Examples are value-at-risk in risk management, advanced derivative pricing models which require performing IT services or on-line retirement planning services were facilitated by both intellectual and information technology innovations.

Figure 1.27 shows the complexity for the risk management function both for the work flows on the business layer and the IT layer, i.e. the network of IT systems. This figure should give an impression why 20-40 percent of the employees work for IT and logistics.

The importance of infrastructure in business can be seen for investment banks (IB). Due to the recent cuts in risk capacity, the increasing costs of risk and regulatory capital and the still present pressure on margins innovation leads to a shift in traditional investment banking: Earnings are more and more obtained from infrastructure banking whereas earnings from traditional IB decrease, see Figure 1.28. Infrastructure banking has at least two meanings. First, banks are offering their trading functionalities to clients. They then can act to using the trading functions of the investment bank without the investment bank using its balance sheet as financial intermediary or they can use the trading function to taylor made the products for their distribution channel. Electronic banking where clients can design taylor-made products is such a case.

The development in stock market trading show up in the change of **speed and fragmentation** in trading: Speed at which investors trade has greatly increased and stock trading has become significantly more fragmented in the major trading places. Two basic questions then are:

- To whom does these developments add value?
- What is the impact on risk for the traders, the investors and the system?

We follow Pagnotta and Philippon (2012). Let us first consider trading **speed**. Market centers have made costly investments in fast computerized trading platforms to reduce order execution and communication latencies. This process has gone beyond stock exchanges to include futures, options, bonds, and currencies. These type of investments started in the U.S. and spread in the last decade worldwide. NYSE increased the speed in cash equity trading from 350 milliseconds 2007 to 105 millisecond in 2009 and to 5 milliseconds at the end of 2009. Johannesburg Stock Exchange increased the speed 2011 by a factor of 400 to 126 microseconds. But speed is not only related to the individual financial center but also between the centers. Spread Networks invested 300 Million USD

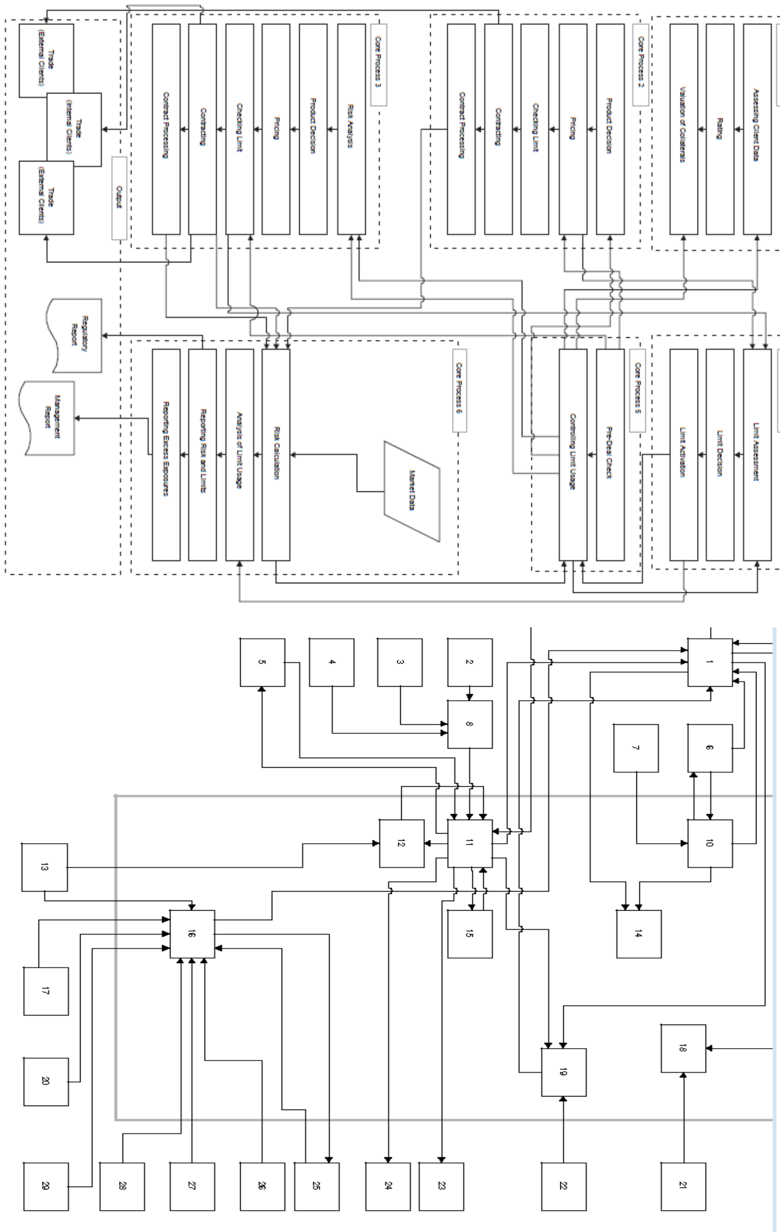


Figure 1.27: Risk management work flows. Upper Panel: Business work flows. On the top level six different functions are defined, the sales functions in work flows 1 and 2, the trading unit (core work flow 3), and the risk management functions in the work flows 4 to 6. The risk management function splits up into the limit management work flows 4 and 5 and a risk measurement work flow 6. The figure shows the information flow, indicated by arrows, between the different functions or units. Lower Panel: The IT system network for the risk management function corresponding to the business work flows. *Source: Leippold and Vanini, 2005.*

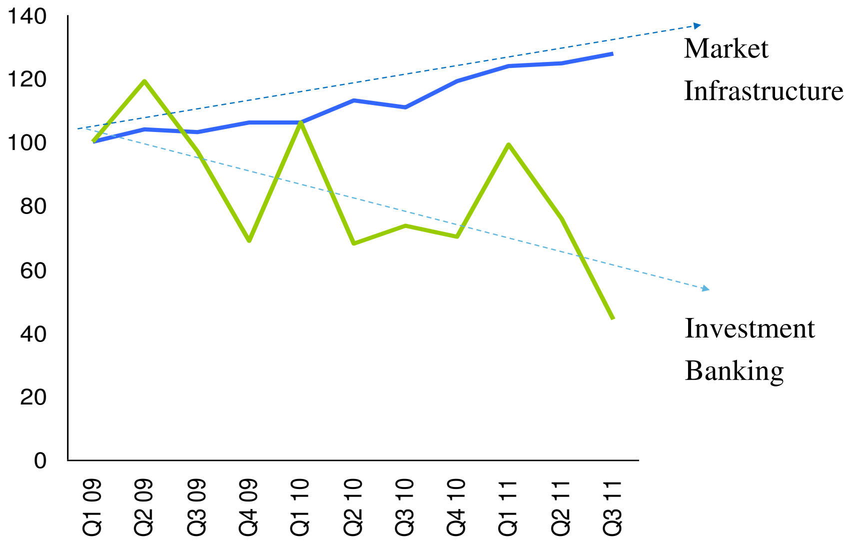


Figure 1.28: Market infrastructure earnings vs. investment banking earnings. The series are indexed to 100 at the beginning of 2009. *Source: Oliver Wyman, 2012.*

in a new fiber optic cable that links New York and Chicago on the straightest route saving 100 miles with respect to existing ones. Speed has an advantage that moral hazard of brokers and traders is no longer possible. The machines are faster than individuals are. This has also a drawback. Suppose that ECB announces at 1200 CET an increase in interest rate. Than machines are able to generate profit from this announcement since they can buy and sell say bonds faster than people can adjust their prices based on news.

Fragmentation means that traditional stock exchangs have lost an still lose market shares to faster such as Chi-X. T The fraction of NYSE-listed stocks traded at the NYSE has decreased from 80% in 2004 to just over 20% in 2009. Most of the lost trading volume has been captured by new entrants. Regulators such as the SEC has enforced fragmentation:
'Mandating the consolidation of order flow in a single venue would create a monopoly and thereby lose the important benefits of competition among markets. The benefits of such competition include incentives for trading centers to create new products, provide high quality trading services that meet the needs of investors, and keep trading fees low.' Source: SEC (2010).

The effects are that large-cap stocks can now be traded in almost 50 venues if one includes bank internal pools, i.e. a bank matches internally supply and demand before routing it in a stock exchange, and other types of markets.

But why do exchanges compete on speed? Is there a relation between the increase in trading speeds and the level of market fragmentation? What are the consequences of these changes? Does fragmentation achieve policy makers' goals? Should investor protection be fostered in the first place? Pagnotta and Philippon (2012) answer these questions based on a microfoundation for how investors value speed in financial markets. They assume that everything else being constant, all investors are better-off by trading faster, but they do not value speed equally. Thus exchanges competing to attract investors can vertically differentiate their intermediation services by catering to different clienteles, relaxing price competition. But why and how investors value trading speed? Preferences need to incorporate heterogeneity to create gains from trade as well as interesting participation decisions among exchanges. First, the flow utility u which an investor derives from holding ψ units of an asset at a given time is

$$u(\psi_t) = (\mu + \sigma\epsilon_t)\psi_t$$

where μ, σ are constant and $\epsilon = \pm 1$ is a random variable. Investors are characterized by the parameters $\sigma \in [0, \bar{\sigma})$ and ϵ. The epsilon shock induce time varying liquidity demands, financing costs or hedging demand while σ measure the size of the shock. The value function of a specific investor at time t is equal to the sum of the expected value of the above cumulated utility up to a random time T, where the investor makes the next market contact, and an expression, which is given by the difference of the value function starting at time T and the trading of the asset at time T at the price p_T.

In their model speed allows investors to realize a larger fraction of the ex post gains from trade. They then analyze the allocation of investors across trading venues. Venues differ in their trading speeds and compete in prices. They show that investors with high expected volatility attach a higher value to speed. The authors then compare the monopolistic venue equilibrium with the equilibrium when two venues are given. Competition lowers fees and increases investor participation. Faster venues charge a higher price and attract speed-sensitive investors. Finally, they analyze the impact of trading regulations aimed at protecting investors. Pagnotta and Philippon (2012):*'We propose a stylized analysis of this regulation by considering two polar cases. In one case, which we refer to as "free segmentation," any venue can refuse to execute the trades of investors from the other venue. The venues are effectively segmented and trades occur at different prices. The other case corresponds to "price protection." We find that price protection acts as a subsidy for the relatively slow market. At the trading stage, investors in the slow venue enjoy being able to trade with investors from the fast venue. Anticipating this, they are more willing to join the slow venue under price protection than under free segmentation.'*

The interaction of technological advances, competition in the securities exchange industry and market regulations interact with each other affecting the trading landscape, asset prices, investor participation, and, ultimately, social welfare is shown in Figure 1.29.

The authors then endogenize the speed and the market structure. They find that price protection encourages entry and that fragmentation leads to more investment in trading technologies and thus faster trading speeds. Their model provides a consistent interpretation of the U.S. experience that after the implementation of new regulations (Regulation National Market System (RegNMS) of 2005) the new market centers proliferated and trading speed increased rapidly.

When they analyze the welfare implications of entry, speed, and investor protection they find that the market outcome is generally inefficient. Pagnotta and Philippon (2012):*'In the monopoly case, participation is always too low and depends exclusively on the distribution of investors. Allowing for endogenous speed improves welfare, even though the speed chosen by the monopolist may be higher or lower than the one chosen by the planner. In the duopoly case, both entry and speed can be inefficient. Regarding entry, there is the usual tension between business stealing on the one hand, and competition and product diversity on the other. Entry typically improves welfare, but it can be excessive if entry costs are relatively high. Regarding speed choices, we find a fairly clear and intuitive condition: Allowing venues to compete on speed improves welfare if the default available speed is relatively low (e.g., purely human-based trading) but decreases welfare once the default speed reaches a certain threshold.'*

The discussion so far left some question open. First, the terms electronic -, algorithmic - (AT) and high frequency tradingHigh Frequency Trading (HFT) are often used interchangeably although they describe different trading behavior. Electronic trading

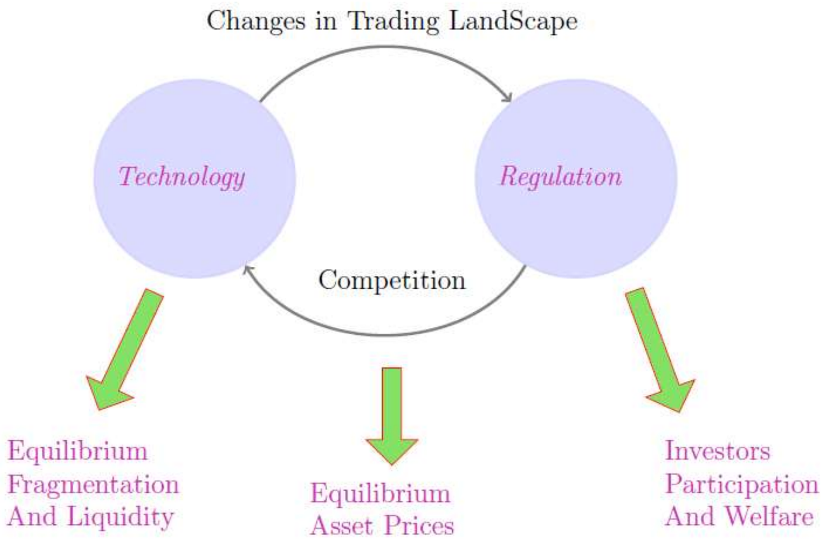


Figure 1.29: Security exchange industry evolution and aggregate outcomes.*Source: Pagnotta and Philippon, 2012.*

simply refers to the ability to transmit orders electronically. AT is a term which does not necessarily implies the aspect of speed. Originally, AT was designed to trade large order by reducing their market impact, i.e. the algorithm were used optimize trade execution. AT is defined as a set of rules which can determine the time, quantity, price, order of routing, monitoring of different venues, etc., see Chlistalla (2011), Brogaard (2010). HFT is a subset of algorithmic trading where a large number of small sized orders are sent into the market at high speed, with round-trip execution times measured in microseconds (Brogaard, 2010). Empirical evidence reveals that the average U.S. stock is held for 22 seconds. While AT are strategies which use technology, HFT is not a strategy per se but a technology, i.e. a form to implement more efficiently trading strategies. Two strategies are (see Chlistalla (2011), Brogaard (2010) for more details):

- **Liquidity-providing strategies** mimic the traditional role of market makers, but unlike traditional market makers, they have no formal market making obligation. These strategies involve making a two-sided market aiming at profiting by earning the bid-ask spread.

- Using **Statistical arbitrage strategies** traders seek profit from imbalances in prices between cross-border or domestic marketplaces for example or between futures on an index and the underlying stocks.

HFT are mainly proprietary trader, i.e. they use their own capital and put it at risk. A. Sussman et al. (2010) report that 48 percent of U.S. equity HFT is due to independent proprietary firms, 46 percent of Broker-Dealer proprietary desks and the rest to Hedge Funds. A main characteristic of HFT is the low latency, i.e. the time between the entry of an order and the execution. Another feature of HFT is the real time analysis of information to produce automatic trading decisions. HFT generate a hugh number of orders whereas a hugh fraction of them is cancelled. Contrary to AT, HFT close their open positions at the end of the day. They do not start the next trading day with unhedged positions. The main reasons for AT are costs, anonymity and trader productivity (they add up to 56 percent). Speed is estimated to be the reason in about 11 percent of all trades.

Jovanovic and Menkveld (2010) suggest that HFT added liquidity to the markets, reduced spreads and reduced arbitrage opportunities across different markets. Although there is no proof of a negative liquidity impacts of HFT, some structural weakness exists for sure. HFT are not obliged to provide liquidity as ordinary market makers. This may lead to an outflow of liquidity if markets are under stress since many HFT programs react in the same way. HFT does not contribute significantly to market depth due to marginal size of their quotes. The low latency and the large cancellation rate of orders make orders of HFT almost not accessible to market participants. Following Brogaard (2010) analyzing a large data set of NASDAQ:

- HFTs add to the price formation process since they follow a price reversal strategy.

- HFTs do not engage in forbidden front-running.

- HFTs provide the best bid and offer quotes for a significant portion of the trading day, but only around a quarter of the book depth and reduce their supply of liquidity only moderately as volatility increases.

- HFTs use fewer strategies than non-HFTs. This is a risk source for market stability if strong market movements occur.

These observations enter the currently ongoing debate about how meaningful HFT is. While electronic trading allowed retail investors to have access to the markets as quick as professionals, under HFT asymmetries are introduced in markets: Special arrangements give preference to specific needs of HFTs. The European Commission intends to subject HFT to MIFID requirements and to supervision by a competent authority. '*The Commission proposes to make sure that all persons involved in HFT above a minimum quantitative threshold are obliged to be full regulatory oversight and to a number of organizational prerequisites such as risk management obligations and capital requirements. In addition, the Commission intends to introduce amendments to MIFID related to the provision of liquidity by HFTs: According to these plans, operators of regulated markets would be required to ensure that a HFT firm continues to provide liquidity on an ongoing basis subject to conditions similar to those applicable to market makers, if it executes a significant number of trades in a certain instrument. In terms of order persistence and tick sizes, operators of regulated markets may be required to ensure that orders remain in the order book for a minimum period before being cancelled – or alternatively to ensure that the ratio of orders to transactions executed by any given participant would not exceed a specified level. Implementing measures could further specify minimum tick sizes that would generally apply to all trading, not just automated trading.' Source; Chlistalla (2011)*

1.3.1 Point of Sale

Recent public available technologies allow a fundamental reshaping of the **point of sales**, i.e. the interface between the banking institution and their clients. They allow for a closer nearness between banking solutions and the clients. The technologies such as the IPad make it possible that clients explore the products in a dynamic and user friendly way by using their fingers and moving back and forward in the product descriptions according to their needs. This replaces the more static, more one-directional traditional communication between the relationship manager and the client using a paper presentation. Using these technologies banks can get off the risky process to first profile the clients and then to match the profile to the products. The alternative is that clients directly search for their appropriate solution. But this requires that clients wants to get involved in the process. It is clear that a fraction of the population will have no interest to do so - even if they can use the new technologies. This technological changes and regulations at the point of sale (MiFID) lead to conjecture:

Conjecture 1.3.1. *Bank clients will in the future use a mandate or make investment decisions individual responsible using the infrastructure of banks. The traditional form of* *consulting by advisory bears to many risks for the banks.*

Complexity and **risk vs. uncertainty** are of a fundamental behavioral importance at the point of sale:

- Complexity is related to question 'how should the bank define the segmentation of customers?'
- Risk vs. uncertainty is related to the question how people perceive risk or uncertainty and how does one decides optimally.

1.3.1.1 Point of Sale: Complexity

It is a fact that most bank use a segmentation of their private customers according to their wealthiness - retail clients, wealthy clients, private banking clients, high networth clients and ultra high networth clients. Different institutions define different threshold levels for the different classes and different assets are considered in the classification (cash, securities, pension fund capital, real estate, etc.). Why do banks use a **segmentation**? The first argument is the return-cost trade off. From a client where a bank can earn USD 1'000 per annum in terms of fees the costs in the advisory should be appropriate. Second, not all relationship managers are equally educated and experienced. This makes a matching of less experienced advisors to less demanding clients meaningful. Two requirements for segmentation are its operability and the adequateness for the clients. As discussed, **wealth** is the key segmentation parameter. But if one requires that the client understands the investment products, unless he has not delegated the decision or uses the banking infrastructure for transaction only, the single wealth variable is not able to hedge against possible advisory risks.

Two major dimensions of any banking product are its risk and return characteristic and its complexity. This implies that there can be complex products which possess little risk and on the opposite, highly risky products which are simple to understand. These two dimensions capture preferences (risk and return) and know how (complexity) of the client. While there are many risk and return measures comparatively little is known about how to measure and classify complexity. We follow a information theory approach. The information content of any financial product can be described by a finite number of questions and there answers. For derivatives, a simple question is 'Which role does the issuer of a product plays for you?', more difficult is 'What does capital protection means in capital protected notes?' and even more difficult 'What does the optionality worst of means in a barrier reverse convertible product?'. Assume that the difficulty level of all questions are $K = \{K_1, K_2, K_3, K_4\}$ and that F_a is the set of questions which are necessary to unreveal all information of product a. We assume that questions in class K_i are less complicated than in class K_{i+1}. A questions belongs

- to class K_1 if it can be answered by yes/no, if the answer does not requires a probabilistic judgement and if no dependency assessment is necessary. Examples:

What is an issuer of a product, what are issuer-related risks which matter for the investor?

- K_2 consists of all questions which need a simple probabilistic judgment, i.e. conditional expectations and dependency issues are excluded. Example: In the last 10 years the probability that a coupon between 2 and 3 percent was annually paid was 45 percent.

- K_3 is the set of questions based on conditional probability assessments and on payoffs with barriers. Example: The probability that the product pays a return of 8 percent p.a. at maturity is 70 percent, given that the underlying does not hit during the life time of the product a barrier level of 60 percent.

- A question is in K_4 if it requires the concern of correlation risk or if no probabilistic dependencies need to be understood. Example: The payoff of an investment depends on the performance of a stock basket of 20 S%P 500 stocks: The return is calculated using the positive performance (if any) of 8 best performing stocks plus the negative performance (if any) the 3 worst performing stocks.

This is a raw characterization which serves for illustrative purposes only. The complexity of the product C^a is defined by

$$C^a(s) = \sum_{i \in K} \sum_{j \in F_a} f(s_{ij})$$

with s_{ij} the difficulty degree of question j in class i and f a valuation function. y f is a convex function in the complexity parameter i. Convexity takes into account that the difficulty degree between the classes K_i is not growing linearly. We require that the complexity measure C^a satisfies:

$$C^a(\lambda s) = \sum_{i \in K} \sum_{j \in F_a} \lambda^i f(s_{ij}) \ , \quad \lambda > 1.$$

The concept is applicable not to single products only but also to portfolios. We therefore require:

$$C^a(s) + C^b(s) = \sum_{i \in K} \left(\sum_{j \in F_a} \lambda^i f(s_{ij}) + \sum_{j \in F_b} \lambda^i f(s_{ij}) - \sum_{j \in F_a \cap F_b} \lambda^i f(s_{ij}) \right) \ ,$$

i.e. the complexity of two products is equal to the sum of their single complexities minus all double counting. As a normalization, we require:

$$C^a(0) = 0 \ .$$

As an example, we consider using the four categories K_j defined above, $f(s_{ij}) = s^i$ and $s = 5$. Three structured products are analyzed:

- A knock-out call option (KO call) with barrier equal to strike.
- A single underlying barrier reverse (BRC) convertible.
- A capital protected note (CapProtect) with coupon payment linked to DJ Eurostoxx 50 and capital protection of 100 percent provided by a AAA-issuer.

We calculate the complexity value of these three products net of all questions which appear in all products, i.e. the complexity value is the value which we have to add to the same unknown and irrelevant basis complexity figure. We get:

$$C^{\text{KO call}} = 305 \ , \ C^{\text{BRC}} = 580 \ , \ C^{\text{CapProtect}} = 1'215 \ .$$

The capital protected note requires the understanding of correlation in an index and the impact of correlation on the likelihood to receive a coupon. This drives the large complexity number. If we define the numerical values for the classes $\mathbf{K}_i$, i.e. $\mathbf{K}_1 = [0, 200]$ for example, the results of the complexity analysis can be compared with the risk of the products. Such intervals are defined as follows:

- The difficulty degree of the questions and the number of questions define whether a product belongs to an interval $\mathbf{K}_i$: The complexity of many questions, each of them of the simple type K_1, can be as high as the complexity of a single, very complicated question which requires the capability of sophisticated statistical reasoning.
- To define these classes requires that the bank uses self-assessments to define the intervals and testing with the function f such that different products are finally classified properly in the scheme of intervals: If the intervals are too small, less complex products are classified in too complex category and vice versa, if the intervals are to broad, complex products are show a too low complex measure.
- The different $\mathbf{K}_i$ then define education levels not only for the client but also for the relation ship managers. Juniors, say can consult only products of the class $\mathbf{K}_1, \mathbf{K}_2$ and after years of experience and education they can gain access to more complex products.

Figure 1.30 shows these two dimensions, where risk is measured as Value-at-Risk (VaR) and the six risk categories are those defined by the Swiss Derivative Association (SVSP). It follows, that the product with the lowest risk figure (CapProtect) is the most complex one and the product with the highest risk, the warrant call, is the less complex one. This negative relation between risk and complexity of investment products in the example shows that differences in preferences (here risk) and differences in understanding (here complexity) are not simply related, i.e. the riskier a product, the more complex is the product. If this would be the case, suitability would become straightforward in the sense that one could sell risky products to experts and to lay individuals only low risk products can be offered. But in the example, the bank has to put the highest efforts into the lowest risk product to be suitable for only few people, those which have a high capability to understand complex issues, can buy a capital protected note. This is an absurd situation.

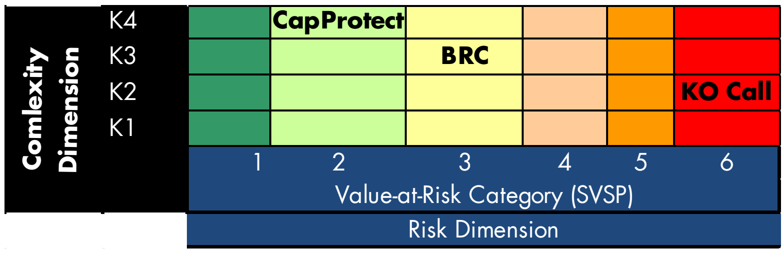


Figure 1.30: Complexity and risk dimension for investment products. SVSP stands for the Swiss Derivative Association.

A bank will therefore decide that full understanding a product with a 100 percent capital guarantee is not relevant - it suffices that the client understands under which condition the capital guarantee is lost (default of issuer). On the other hand, suppose that a product is both of maximum complexity and maximum risk. Then, the bank should ask (i) whether there are enough educated relationship managers able to understand and explain the product and (ii) how many clients the bank has in this sector. This may well lead to the conclusion that the product is withdrawn because consulting risks and the few number of clients are not acceptable. This shows that using complexity and risk structures the point of sales: If forces the bank to think about its own capacities and the capacities of the client base in a concrete way, i.e. the product offering of the institution.

1.3.1.2 Point of Sale: Risk and Uncertainty

Risk and **uncertainty** are two different concepts, leading to different behavior and with a different neurological foundation.[37] According to Knight (1921) risk refers to situation of perfect knowledge about the probabilities of all outcomes for all alternatives. This makes it possible to calculate the optimal. Uncertainty refers to situations where the probabilities are unknown or unknowable. Robust decision making is a specific setting where one calculates optimal solutions assuming that a set of probabilities can influence the state variables and where nature chooses the worst probability, see Section 1.8 for a specific model in intertemporal decision making. Knight perceived that uncertainty

[37]Gigerenzer and Volz (2012), Gigerenzer and Goldstein (2011), Kahneman (2011), Knight (1921), Savage (1954), Binmore (2009).

may afford opportunities for profit that do not exist in situations where risks can be calculated.

Savage (1954) introduced the term '**small worlds**' for situations of perfect knowledge where all relevant alternatives, their consequences, and their probabilities are known for certain. According to him, these are the worlds in which Bayesian theory provides the best answer. Examples are lotteries and roulette. Small worlds need to be distinguished from '**large worlds**,' where part of the relevant information is unknown or must be estimated from small samples, or the future is uncertain. Examples are decisions about when to plan a picnic, whom to marry, and how to raise your kids.

Decision making under uncertainty is what our brain does most of the time, while situations of known risk are relatively rare and found mostly in gambling. Savage argued that applying Bayesian theory to decisions in large uncertain worlds would be utterly ridiculous' because there is no way to know all alternatives, consequences, and probabilities. As a consequence, the brain needs strategies beyond Bayes' rule to succeed in an uncertain social and physical environment.

The best solution in a world of risk is generally not the best one in a world of uncertainty. If for risk value-based statistical thinking (e.g., Bayesian probability updating plus utilities) is sufficient for making good decisions, provided that the problem is computationally tractable then for uncertainty statistical thinking is no longer sufficient but heuristic thinking is required. The recent financial crises illustrate that statistical tools for estimating risk, Bayesian or otherwise, failed consistently in the real, uncertain world of finance (Taleb, 2010). They are optimal when risks are known, but not in a world of uncertainty. Applying normative theories of risk to uncertain worlds can in fact lead to disasters. With respect to the financial crash of 2008, Stiglitz (2010) noted: *'It simply wasn't true that a world with almost perfect information was very similar to one in which there was perfect information'*. In sum, norms derived from assuming known risks do not simply generalize to norms under uncertainty.

Consider how to allocate money to N assets is Markowitz's mean-variance model. Like all optimizing theories, it assumes a small world with perfect knowledge about the relevant parameters. Is this theory also optimal in the real, uncertain world of financial investment, where parameter values are not known for certain but need to be estimated? De Miguel et al. (2009) compared the mean-variance model with a heuristic called $1/N$, or equality heuristic. The heuristic simply allocates money to N assets equally. The result was that $1/N$ consistently performed better in out-of-sample prediction (an elementary form of uncertainty). Cross-validation is a prime example of out-of-sample prediction: the data is divided into two complementary subsets: the in-sample data set, which is used for fitting the parameters of the competing models and an out-of-sample data set, which is used for testing how well the models predict (see also below). Note that in data fitting, that is, when all data are known, the optimizing model always wins, but not in prediction. None of 12 other optimization models, Bayesian or otherwise, could consistently predict better than the simple heuristic. This result contradicts the widespread view that heuristics are always second best to logic and statistical optimization models. This view makes no distinction between risk and uncertainty. Researchers in this tradition have evaluated people's reliance on 1/N negatively and attributed it to their cognitive limitations. However, ignoring part of the information is what makes heuristics robust for the unknown future, whereas by trying to integrate all information and estimate the weights, complex strategies such as the mean-variance portfolio suffer from overfitting the past. The mathematically sophisticated reader who wants to understand why and when simple heuristics can be more accurate than complex statistical methods will find an answer in the bias-variance dilemma. The fact that simple heuristics often outperform "optimization" models in situations of uncertainty has been demonstrated many times over (Gigerenzer et al., 2011). In order to deal with an uncertain world, the brain relies on an adaptive toolbox of heuristics. Accordingly, intelligence is defined as the degree of knowing in which situation to use which heuristic. The scientific study of this normative question is called the study of the ecological rationality of a heuristic. For instance, $1/N$ tends to outperform mean-variance optimization in situations where predictive uncertainty is high (stocks are hard to predict), the number of options N is large (the optimization models have to estimate more parameters which leads to more error), and the sample size is relatively small. In uncertain worlds with these features, $1/N$ can be expected to be both faster and more accurate than the mean-variance optimization. When would mean-variance outperform $1/N$? De Miguel et al. (2009) estimated that with 50 assets, one would need some 500 years of stock data before the optimization model is profitable. Humans rely on the $1/N$ heuristic not only for financial investment. In many situations, fairness and justice are achieved by distributing resources equally. Our normative argument has fundamental consequences for the neuroscience of decision making: Claims that the rational brain always works by Bayesian calculations are founded on the assumption that what is rational in a world of risk is also rational in an uncertain world - the world our brain has to deal with most of the time. These claims are also incompatible with three well-known restrictions: Bayesian optimization is not feasible if (i) the choice alternatives are not known for sure, (ii) the mind has more than one goal, and (iii) even if all alternatives were known and the mind had only one goal, the calculations can quickly become computationally intractable, that is, no mind can actually perform them in a lifetime (Gigerenzer, 2004). Bayesian inference works in small worlds where there are reliable data for probabilities and only a few alternatives and cues.

We argued that what is optimal in a world of risk is typically not the best in a world of uncertainty. Consequently, an adapted brain relies on different processes according to the situation. When faced with risk, using heuristics is of little value, unless the computations become too difficult. When faced with uncertainty, using logic and statistics is of little value, unless the part of the problem that is known is being calculated. In an uncertain world, there is broad experimental evidence that humans and other animals rely on a toolbox of heuristics. These are based on evolved and learned core capacities and include (see Gigerenzer and Gaissmaier, 2011):

- Recognition-based heuristics: Recognition heuristic; fluency heuristic.

- Equality-based heuristics: $1/N$.
- One-good-reason heuristics: take-the-best.
- Social heuristics: tit-for-tat, imitate-the-majority.

Consider a simple heuristic that humans and other animals use to make inferences about an uncertain world: Recognition heuristic (RH): If one of two objects is recognized and the other is not, then infer that the recognized object has the higher value with respect to the criterion. Goldstein and Gigerenzer, 2011: *'Consider the question whether Milan or Modena has more inhabitants. If one has heard of Milan but not of Modena, the inference is that Milan is the larger city. Note that the RH requires semi-ignorance to be applicable, meaning that if one has heard of both (or neither) objects, it will not be effective. Experimental studies indicate that a large proportion of subjects rely on it in uncertain situations, such as when predicting which tennis player will win in Wimbledon or which political candidate to vote for, and by animals when choosing food.'* Using fMRI, researchers find evidence that RH-based decision processes go beyond automatically choosing the recognized alternative and are guided by judgments about the ecological rationality of the RH, as reflected by activation in anterior medial prefrontal cortex.

1.3.1.3 Model Risk as a Type of Uncertainty

When consider a specific form of **model risk** :[38] **Uncertainty** about the risky assets price dynamics. We assume that ...

- ... all agents are rational.
- ... all states of the world are known.
- ... the agents do not know which probability distribution drives the risky asset dynamics. In this sense there is uncertainty in decision making.

Consider a risk less asset B and single risky asset S. The risky asset S dynamics is driven by a constant drift plus a noise or random term. In most traditional models one assumes that the probability distribution P is **known with certainty** - this is the zero model risk situation. Model risk is inserted as follows: We assume that the decision maker does not know P - other probability laws could as well drive the asset price dynamics. Let P' be a different law. P and P' can be very different but they have to possess the property: An impossible price paths under P is also impossible under P' and vice versa (the probabilities are called equivalent). If under P the path that the stock price of say Google doubles within one year is 5 percent, then under P' the same event with a probability of 50 percent or 0.05 percent is allowed - but not with 0 percent.

[38] We refer to Trojani and Vanini (2002) for formal details in the following discussion. Literature concerned with ambiguity is Brenner and Izhakian (2011), Dow and Werlang (1992), Epstein and Wang (1994), Cao et al. (2005) and Ui (2011), Izhakian (2012a), Schmeidler' (1989).

Why is equivalence of probabilities or priors required? The most convincing answer is that in this setup powerful tools from probability exist to analyze these problems.

A second property bounds P' from to be 'too distant from P' - we limit the extent of uncertainty. How do we define a measure stick for the distance between probability laws? This is not trivial since the set of probability laws is not a vector space where the usual measure stick applies. The quantity used is **relative entropy** , i.e. a measure of the discrepancy of the two probabilities. This can be interpreted as the expected surprise ex-

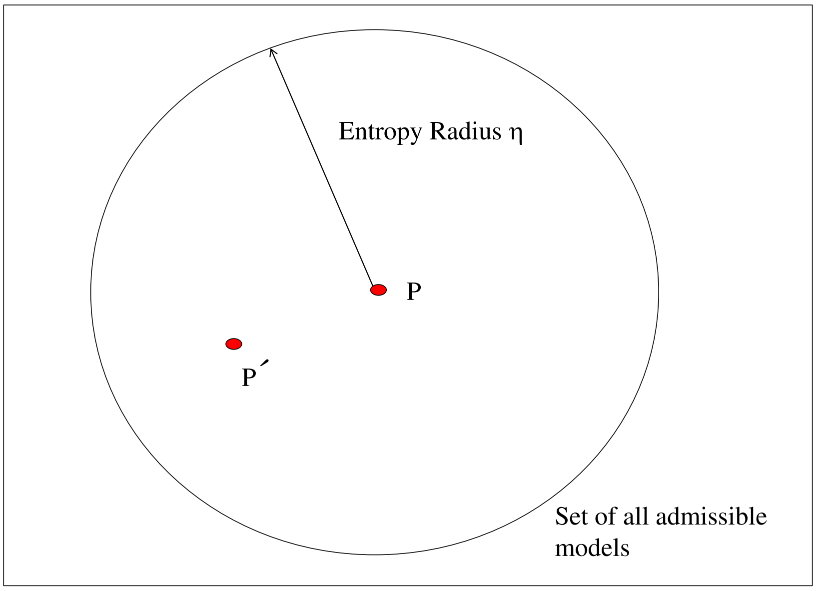


Figure 1.31: Set of model misspecification and relative entropy.

perienced when believing that P describes the model dynamics and being informed that in fact these are described by P'. Relative entropy is zero, if $P = P'$, see Figure 1.31. We interpret the entropy measure as the marginal rate of change with which expected surprises are experienced when we are continuously informed over time about the underlying data generating mechanism. How does an agent makes his optimal decision in such a uncertainty-related setup? As a first step the entropy radius is bounded by a number η. η is the largest entropy distance for which a model misspecification is seen as relevant by the agent - doubling of the Google stock price in one year is acceptable but quadrupling is not considered to be possible. Then the investor chooses his optimal investment strategy and consumption to maximize expected utility under the budget constraint. This is also deformed due to the model risk of the asset price dynamics entering the constraint. So far, the decision problem is over the set of all different probabilities within the entropy radius. Which one should the agent consider? The worst case is to assume that a second player - nature - chooses the probability which is admissible with the worst consequences for the investor. This is called **a robust optimal decision problem** . In this sense, the optimization problems leads to optimal decisions by the agent where the model misspecification result deviates maximally from the reference one P. Carrying out the analysis for an isoelastic current utility of consumption $c^a/a, a \in (0,1)$, the optimal robust consumption and investment rules satisfy:

- A lower variance of equilibrium consumption and a lower variance of expected growth of equilibrium optimal aggregate consumption compared to the non-robust problem follows.
- Lower investments in the risky assets, i.e. a larger fraction of wealth is invested in the risk free asset compared to the non-robust case.
- A decomposition of the market price in the usual market price of risk part plus a an extra equilibrium reward for risk that arises because of a possible misspecification of the given reference model for asset prices. The usual consumption based market price of risk is of the form market price of risk times risk tolerance (i.e. the parameter a matters). The market price of risk equals excess return over the risk free rate divided by volatility.

This description of a basic model of a formal model risk is generalized in several directions.

- Equilibrium analysis, i.e. the impact of model risk on endogenous asset prices is considered
- Learning, i.e. agents learn about uncertainty.

We do not solve this model analytically since this would take us to far away. But we solve the non-robust version of this model and then simply state the robust solution.

The non-robust problem in the discussion above is Merton's optimal consumption and investment problem. We discuss this model next which is a classic example of an **intellectual innovation** . In order of not getting lost in the mathematical formalism we first consider the structure of intertemporal decision making. An agent chooses a decision (vector) variable c, which is a function of time. The decision can be a consumption choice, a choice of portfolio, a cost function, a salary scheme or a combination of them. This variable is chosen such that the objective function

$$E^P[\int_T u(t,c,W)dt]$$

is optimized with $u(t,c,W)$ the instantaneous utility function and W the state variable(s). Instantaneous utility $u(t,c,W)$ depends on the date, the choice at c at this date and the state variable W. The (vector) state variable W describes the evolution of states such as wealth, human capital, risk measures such as Value-at-risk. The set T defines the time horizon of optimal decision making. P is the probability which the agent assumes to account for randomness in the state variable - i.e. if W_t is an asset price today, the price tomorrow is not know with certainty. This optimization is done under the restriction that $(dW_t, c_t) \in A_t$, i.e. that the state and decision variable assume feasible values. For the state variables such a feasibility condition is the budget restriction, i.e. a dynamic relation for the state variable. For the choice variable consumption feasibility for example means that consumption is never negative. Solving the optimization problem provides an optimal choice function c^*.[39] This function depends on:

- The agents preferences u.
- The probability law P
- The assumptions about the constraints set A_t.

These are all sources of model risk. Preferences can be misspecified, the probability law over uncertainty can be distorted, the dynamics of the state vector flawed: If a state vector turns out to jump over time but the dynamics does not allow for jumps or if it turns out that it is very difficult to estimate the state vector parameters. This is prominent in portfolio theory where the necessary estimates of expected returns and correlations are plagued by estimation risk.

Whereas in the discussion so far the concern is about misspecification of the randomness of the state variable dynamics there are situations where also consequences or the states themselves are uncertain. The A-influenza is such an example where not only the probability of the spreading disease are unknown but also the states or consequences in terms of number of deaths is largely unknown.

We formalize Merton's model which serves as a basic model for many inter temporal finance optimal decision making models.

We start with an individual which at time t_0 wants to maximize utility in a period $[t_0, T]$ by making **continuous** decisions c_t. One assumes that the dynamics of the state variable follows a **diffusion process**:

$$dW = g(t, c, W)dt + \sigma(t, c, W)dB \ .$$

The first term is the drift, the second one the noise term. The drift defines the evolution of say wealth if a small amount of time dt elapses. The drift coefficient function g is setup in a general form, i.e. the agent can control it through the variable c and it can be state dependent. The same remarks apply to the **volatility function** σ. The expression dB represents an increment of Brownian motion, i.e. it can thought about a difference between two random variables, closely apart in time, with the properties: First, $(dB)^2 = dt$. That is displacement in space scales as the square root of time. Second, the expected value $E(dB) = 0$. We consider this **basic building block** for financial modelling in continuous time later on in more details. We just note that this stochastic process, i.e. for all t we have random variable B_t (we skipped time index

[39] Only a few models can be solved analytically.

above), has a long history starting with the biologist Brown, the mathematician Batchlier and Einstein in 1905, Samuelson, Black and Scholes in the 60s and 70s of last century.

If J is the value function, the agent solves

$$\begin{aligned} J(t_0, w_0) &= \max_{c,\omega} E\left[\int_{t_0}^{T} u(t,c,W)dt + f(W(T),T)\right] \\ dW_t &= g(t,c,W,\omega)dt + \sigma(t,c,W,\omega)dB_t \ , \ W(t_0) = w_0 \ . \end{aligned} \tag{1.9}$$

The function f defines terminal utility, w_0 is the initial wealth level and ω is the fraction of wealth invested in the risky assets. This is the second choice variable, the optimal portfolio choice, besides the consumption stream. How do we solve this optimization problem in continuous time? The problem can be solved analytically only in some specific cases, i.e. for some utility functions (log, HARA, CARA, quadratic) and simple functions g and σ. Therefore a large and impressive body of numerical methods emerged in the last decades. We use the **Principle of Optimality** which was developed by Richard Bellman defining the method of dynamic programming.[40] The principle states that an optimal path c_t^* has the property that whatever the initial state conditions and control values of c over some initial period, the control or decision variable c over the remaining period must be optimal for the remaining problem, with the state resulting from the early decisions considered as the initial condition. We show in Appendix 7.1 how the optimization problem is formally solved for the specification:

$$u(c) = c^a/a \ , \ 0 < a < 1 \ , \ f(\cdot) = 0$$

a Geometric Brownian motion dynamics for the risky asset with drift μ and volatility σ and a risk less asset with return r.

The optimal decision and value are:

$$V(W) = \alpha^* W^a \quad , c^* = W(a\alpha^*)^{\frac{1}{a-1}} \ , \ \omega^* = \frac{\mu - r}{\sigma^2}\frac{1}{1-a} \ .$$

with α^* a W-independent constant (see the Appendix). This shows that it is optimal for the individual to invest in the risky asset (and therefore also in the risk free asset) independent on the state variable. Optimal investment is proportional to the market price of risk $\frac{\mu-r}{\sigma^2}$ and relative risk aversion $\frac{1}{1-a}$, i.e. a product of market characteristics and individual preferences. This observation is common to most inter temporal consumption-investment problems. Optimal consumption is proportional to wealth, i.e. state dependent. In more complicated models with model risk the optimal investment variable is no longer state independent. it often becomes dependent on the state variable describing the stochastic opportunity set. This set can be stochastic due to model risk

[40]There are two other solution methods, one is the optimal dynamic control of Pontryagin and the calculus of variations which dates back several hundred years to Lagrange and Euler. Which method to use depends on the problem which one considers.

or changing macro economic circumstances. In the optimal consumption and investment model with model risk discussed above, the optimal decision rules change to:

$$c^* = W(a\alpha^*(\eta))^{\frac{1}{a-1}} \ , \ \omega^* = \frac{\mu - r - \sqrt{2\eta}\sigma}{\sigma^2}\frac{1}{1-a}$$

with α^* a wealth-independent constant (see the Appendix). The uncertainty parameter η therefore enters in the market price of risk and the optimal consumption path. This shows that an investor facing model risk will invest less in the risky asset compared to the investor without model risk. There are various extensions of this basic model: Models with learning, multiple assets, more realistic constraints. We refer to the literature for the formal methods, see Anderson (2005), Hansen and Sargent (2001), Hansen and Sargent (2008), Trojani and Vanini (2002). Using the expression for the optimal consumption and investment strategy the price dynamics of the risky asset and the interest rate of the risk less investment follow from the market clearing condition in both markets.

Although the above model and its extensions are based on first economic principles they are hardly used in practice. Uncertainty and risk are considered in a different way. First, uncertainty is not defined in the above way that one does not know which probability laws applies but one knows the set of all possible laws. Uncertainty in practice is not a well defined statistical concept. Uncertainty is often related to **bounded rationality** and lack of knowledge about future states. Therefore, to deal with uncertainty and risk in practice is much less related to sophisticated formal modelling but more to reasoning. As an example consider a bank setting a price p for a product. Uncertainty arises if the bank consider two other prices: p_m the market price of the product and p^* the price which the bank assumes should be paid for the product. The following alternatives are possible:

$$\begin{aligned}
p &< p_m < p^* \\
p &< p^* < p_m \\
p_m &< p < p^* \\
p_m &< p^* < p \\
p^* &< p_m < p \\
p^* &< p < p_m \ .
\end{aligned}$$

Why should a bank assume that there is a price p^* different from market prices p_m? First, the bank should not assume this but think about whether this is in their opinion the case for a specific product. That is, the bank has to use their experience and analytical skills to discuss whether the market misprices a product. The six alternatives define different risks and market shares for the bank. The simple answer to set $p = p_M = p^*$ is only optimal if there is no uncertainty at all. In all other cases, such a choice is not optimal: The bank misses opportunities if they can figure out the right inequality relation and runs large risks if they choose the wrong one. If it turns out ex post that market prices were too high, the winning situation for the bank is $p^* < p < p_m$: the bank is on the save side from the risk perspective and earns a rent compared to all other market participant. To arrive at such a decision $p^* < p < p_m$ first requires that market misprice financial instruments and that the bank under consideration has three capacities: First, the bank needs the experience or analytical strength to identify the mispricing. Second, they need to be that efficient and effective that they can offer at lower than market prices. Finally, they need the courage to believe in the view and to act in this respect.

1.3.2 Social Networks

Social networks are mainly interesting for the bank in two respects: Outsorcing of costs and increased quality of client services and preferential attachment of savings to loans. It is well known that agents in social networks are motivated to share their knowledge and to help other people in the network. It is also known that some agents in a social network, which are independent of the banking institution, know more about what other bank clients do not understand than the bankers. They offer free, better and often faster advice than the banking professionals. This defines the business case to setup a social network where some costs of the client relationship management are outsourced with non-decreasing quality of consulting. There are examples of large firms (Swisscom) which implemented such a model and were both costs and quality evolved in the desired direction. But from a **risk perspective** what is beneficial today may be harmful tomorrow. Since the agents in the external network are not banks employees one cannot dismiss them if they cause problems. Furthermore, the opinion makers in the network have many followers, i.e. if a bank is in trouble with an opinion maker the bank is in trouble with a large fraction of the network. The cause of such risks are differences in values: A priori it is not clear why the different values of a bank and of a social network community should not contain conflict potential. Even further, the way of communication in the network has to be transparent to everybody. This leads to a conclusion, that in case of severe problems the only risk management action is to shut down the network. But this leads to a disruption between in the client-bank relationship and triggers severe operational risks. The identification, valuation and management of the risks between a social network which operates in the point of sales interface of a bank and the business continuity management of the bank is still in a state of infancy. It is not clear whether there exist intrinsic motivation such that a network is formed which helps bank to improve their services.

The second aspect of preferential attachment is already implemented in practice: There are banks and networks or networks which have not a banking status where money invested is linked to loans. Basically, a network platform allows borrowers and lenders to post their needs and preferences. The bank needs matches them. An interesting issue is who bears credit and liquidity risk. Consider an individual which is willing to hand its investment to a home owner. The home owner pays interest to the investor. What happens if the home owner defaults - is it the bank which faces the credit risk or the investor? Another issue is compliance. Many financial s institutions have been reluctant to use social media tools because of the stringent compliance and regulations that govern the sector. This restricts the use of social media as a new channel for traditional business.

We consider **Crowd Funding** as an example. With Crowd Funding, people which are looking for money for projects are brought together over the Internet with investors. This idea from the U.S. has recently extended to the financing of startups or SMEs (small and mid sized enterprizes). In Europe at present (2012) 37 platform attack traditional banks in different areas from payments to Crowd Finance. In some jurisdictions, such as Germany, public offerings require a prospectus. Such a prospectus has to inform about all types of risks and needs approval by the Bafin (the German regulatory authority). This rule does not applies if the offering is lower than 100'000 EUR. Hence most providers restrict their business to low capital amounts. Crowd Funding is not risk free. UK's Financial Services Authority (FSA) warns investors to naively participate in Crowd Funding to finance over the internet. *'Many ways to crowdfunding are highly risky, complex, offer no guaranteed return, rarely dividends, and if anything, it takes a long time until investors see results'*, criticizes the FSA (2012). Furthermore, investments are very illiquid and no secondary market exists. Almost none internet page is regulated or even controlled, i.e. there is no client protection. Some platform realize that circumventing regulation will not add value to their business but limit their success. Recently the platform Seedrs, a Crowd Funding platform for startups, is authorized and regulated by the Financial Services Authority. It took Seedrs two years to bring their business model in line with the British and European regulatory requirements and the FSA then needed another year to provide the approval. The platform offers the two classical streams - invest in startups and raise startup capital.

1.4 Who Innovates, Innovation Life Cycle

Boot and Thakor (1997) model how different institutional structures might lead to different levels of innovation. They find that innovation would be lower in a universal banking system—especially one with substantial market concentration—than in one in which commercial and investment banking were functionally separated. Essentially, greater competition among these private parties leads to increased innovation. There is mixed empirical academic evidence about which type of institution innovates and who follows the innovation. But from a practitioners view banks with larger market shares will tend to innovate, as will banks whose clients are more sticky. Innovators earn higher market shares than followers, even though imitation is rapid. Innovation has a difficult to measure impact on reputation. Some investment product innovations face the life cycle shown in Figure 3.28.

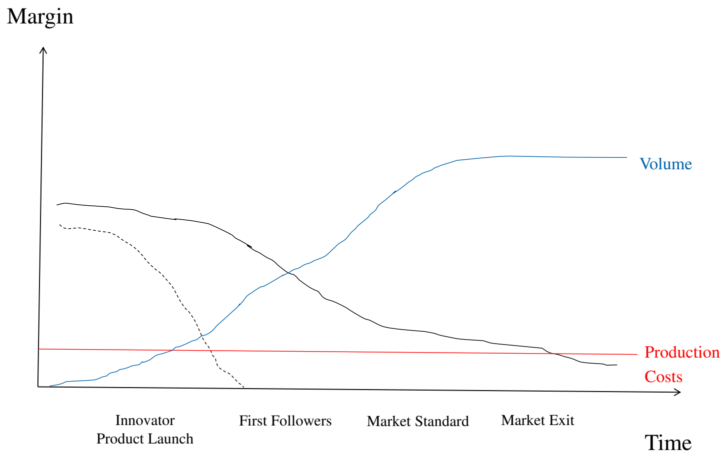


Figure 1.32: Life cycle for investment products.

The first player is the innovator. If the innovation is successful the highest margin due to monopolistic power follows. If innovation fails, the dashed line, the product vanishes. This is more the rule than the exception for financial innovation. But failures provide information for subsequent innovations. The next step of successful innovation is the entrance of the first market competitors which offer the same product (Followers). To copy a financial product is a matter of few days for simple structures and for more complicated ones the advantage of the innovator should not be expected to be longer than roughly a half a year. This possibility to copy products that fast is due to the **absence of patent rights** and the extensive intellectual capacities within the investment banks. We then enter in a liquid market case. Whether the product is produced by all banks in the market or whether some buy it from other ones and serve as a distribution channel to their clients, depends on the product's complexity, the bank's know how and risk appetite. In the last period of the life cycle efficient production matters, i.e. the production costs become important. Those banks with inefficient work flows stop selling the product. Production costs have several components. There are direct costs for the work force, IT costs, costs for distribution, marketing, education, regulatory and economic capital costs.

Understanding the innovation dynamics has been a long-standing research topic. Merton (1992) characterizes the dynamics of innovation in the financial service world using a metaphor of 'financial innovation spiral' in which one innovation begets the next. We see the spiral when we consider that the trading of standardized exchange-traded products facilitates the creation of custom-designed OTC products, which in turn stimulates even greater trading, lowering transaction costs and making possible even more new products.

In Merton Miller's 1986 view on financial innovation, the period from the mid- 1960s to mid-1980s was a unique one in American financial history. Contrary to his belief about future innovation, financial markets continued to produce a multitude of new products (derivatives, alternative risk transfer products, exchange traded funds, and variants of tax-deductible equity). Tuffano (2002) states: *'A longer view suggests that financial innovation is an ongoing process whereby private parties experiment to try to differentiate their products and services, responding to both sudden and gradual changes in the economy. Surely, innovation ebbs and flows with some periods exhibiting bursts of activity and others witnessing a slackening or even backlash.'*

1.5 The Impact of Financial Innovation on Society

While most authors acknowledge that innovation has both positive and negative impacts on society, their conclusion regarding the net impact of financial innovation reflects a diversity of opinions. Merton (1992) stakes out one side of the argument: *'Financial innovation is viewed as the "engine" driving the financial system towards its goal of improving the performance of what economists call the "real economy.*' He cites the U.S. national mortgage market, the development of international markets for financial derivatives and the growth of the mutual fund and investment industries as examples where innovation has produced enormous social welfare gains. Others take the opposite viewpoint to make the argument that innovation's benefits are less clear: Time and again, business has seized upon a new idea—junk bonds, LBOs[41], derivatives— only to push it far past its sensible application to a seemingly inevitable disaster.

How do we research the question of the net social benefits of innovation? One approach attempts to measure the size of the gains for specific innovations, say the innovations in mortgage markets in the form of securitization. While some find positive evidence, other researchers often from the legal and policy literature find contrary evidence by discussing the costs due to tax evasion, reduced tax revenues, loss of confidence in government and social costs of inequality or inequity. Other arguments against welfare gains are complexity that in turn leads to bad business decisions and social costs or that specific innovations contribute to high levels of market volatility and possibly to market crashes.

Do derivatives have a positive or negative influence on social welfare? Tuffano (2002) states: *'Despite the best intentions of the authors, their studies cannot measure social welfare directly, nor can they benchmark the observed outcomes against those never observed. Furthermore, in light of the innovation spiral (where successful innovations beget others) and the evolutionary process (where many innovations fail), it is exceedingly difficult to identify the boundaries of a particular innovation, if one wanted to measure its*

[41] A leveraged buyout (LBO) occurs when an investor acquires a controlling interest in a company's equity and where a significant percentage of the purchase price is financed through borrowing, i.e. leverage.

costs. It is a hopeless task to measure the ex post impacts innovations. Ex ante views often focus on very specific and narrow aspects of innovation to permit a meaningful discussion.' The existing theoretical models are too stylized and to narrow to allow for general welfare considerations. Duffie and Rahi (1995), summarize a wide range of the literature: *At this early stage, while there are several results providing conditions for the existence of equilibrium with innovation, the available theory has relatively few normative or predictive results. From a spanning point of view, we can guess that there are incentives to set up markets for securities for which there are no close substitutes, and which may be used to hedge substantive risks.* This summary still holds true today. The complexity of the question contrasts with the strengths of the available analytical tools. If we setup a general economic model yet the specification of the individual preferences is a complicated task if one wishes to capture the time-varying opportunity set with and without innovation and possible feedback effects of decision not related to the innovation part and vice versa. The complexity further increases if one aggregates individual preferences and tries to derive properties of the general equilibrium. Reality shows that such a model could only be treated numerically. One might ask, whether the traditional microeconomic approach is in principle well suited to answer such questions or whether not a different approach is needed. A lot of the heated debate about the dark side of financial innovation reflects the shortcomings of traditional analytic tools where moral, ethics and emotions replace formal thinking.

Given the difficulty to value financial innovation for society, one could propose that there is value if an innovation becomes successful and that one puts the efforts in the identification of potential dark sides of innovations.

There is more than evidence that financial innovations is sometimes undertaken to create complexity and exploit the purchaser - CDO Squared are an example. Some emails from investment bankers which became public show that some bankers indeed follow such a client hostile strategy. Paul Volcker said in December 2009 that the biggest innovation in the industry over the past 20 years had been the cash machine. He went on to attack the rise of complex products such as credit default swaps (CDS). *I wish someone would give me one shred of neutral evidence that financial innovation has led to economic growth — one shred of evidence*, said Mr Volcker. Many others made a similar point. Krugman (2007) argues

(T)he innovations of recent years—the alphabet soup of C.D.O.'s and S.I.V.'s, R.M.B.S. and A.B.C.P.—were sold on false pretenses. They were promoted as ways to spread risk, making investment safer. What they did instead—aside from making their creators a lot of money, which they didn't have to repay when it all went bust—was to spread confusion, luring investors into taking on more risk than they realized.

Henderson and Pearson (2011) provide evidence for a particular type of structured equity product.[42] They show that these were overpriced and did not provide any redeeming service to investors. They document 64 issues of SPARQS by Morgan Stanley from June 2001 to the end of 2005 and show that the return on these risky securities was less than the risk free rate. They are able to show that these securities have no advantageous hedging properties, liquidity features or tax advantages that can explain this low return. During the three and a half years they study Morgan Stanley issued about USD 2.2 billion of these securities. Their payoffs were tied to the stock price of major listed companies. They are typically callable after six months and have a maximum maturity of slightly over a year Henderson and Pearson demonstrate that they have a price premium when they are issued of 8 percent compared to an equivalent dynamic trading strategy with exactly the same payoffs. Given the short maturity and interest rates at the time this means their payoff was less than the risk free rate. Since they are positively correlated with major stock indices they do not have any advantageous hedging properties. They are taxed as prepaid terminable forward contracts. If anything this gives them a tax disadvantage rather than advantage. Moreover, they are not particularly liquid. Henderson and Pearson argue investors would have been better off investing in banks' certificates of deposit. Structured equity products became very popular not only in the U.S. but also in Asia and Europe.

Bergstresser (2008) documents that at the peak structured products reached a total outstanding of Euro 4.4 trillion. He considers a much larger sample than Henderson and Pearson consisting of 314,000 individual notes including issues in Asia, and Europe as well as the US. His results are similar. Prior to 2005, these products were overpriced similarly to those considered by Henderson and Pearson, particularly those issued by Goldman Sachs and Unicredit. However, subsequently this overpricing was considerably reduced. There seem to be many occasions where structured equity products were significantly overpriced in order to extract money from investors who did not fully understand the alternatives to what they were buying.

I do not intend to comment on the adequacy of the used methods neither I want to look for a needle in a haystack. When I compare the issuance margin - the 8 percent difference between the fair price and the issuance price - Henderson and Pearson (2011) provide about the 1y products SPARQS with the issuance margin of the business of a large Swiss bank in the last 3 years across all products, then I conclude:

- Innovation has a **cultural** component.
- The **life cycle** of innovation insight applies.

These claims follow from the issuance margin in all structured products of a Swiss bank which ranges in the last three years between 1.15 and 1.32 percent. The value of the issue amount ranged between 2.5 and 3 Billion CHF. All types of structured products were considered except vanilla option (warrants, knock-out warrants). The difference

[42] They are known as Stock Participation Accreting Redemption Quarterly Pay Securities (SPARQS)

between the pricing of the Swiss bank and the U.S. Investment bank shows that two differences interfere. First, during the period where SPARQS were issued this type of product was not a mass type product from the supply side - only few issuers could and indeed offered the products. In such an early stage of a product life cycle margins are higher as we already discussed. Second, the incentives, total compensation for bankers and the way how the bank considers the point of sales are different for people working for a large investment bank compared to bankers working in smaller, more retailed focus institution. These differences make have a cultural component. This triggers different answers to the question 'Which costs do we charge?'.

The vast range of different financial products, types of financial institutions and services raise the motivation to search for a categorization. But all attempts to catalog innovation face some shortcomings. One might ask what the value of such a categorization is and for whom? It is doubtful whether people who innovate need a categorization scheme.

Innovation can mean for example

1. Products (swaps, options, ...),
2. New corporate securities (tier 1 bonds, hybrid capital, ...),
3. Processes (outsourcing industry production, using new transaction processes, ...),
4. Governance (salary system, point of sales, ...).

Although these types of innovations seem independent from each other they often are not. Innovation often affects several types. Truly novel innovations occur very few. Independent of the originality of an innovation, it has two parts. An act of invention is followed by a diffusion of new products, services etc. The attempt to categorize innovations tends to be either uninformative (firms use names to differentiate similar products), not consistent if legal or regulatory definitions are used since innovations often spans between the defined objects (structured products are a debt-like product but they possess characteristics of equity and other assets classes), not manageable if products features are used for categorization (e.g., maturity, redemption provisions, etc.). Academics prefer to characterize products by their function they serve, see the BIS approach. Merton's (1992) functional decomposition identifies six functions delivered by financial systems: (1) moving funds across time and space; (2) the pooling of funds; (3) managing risk; (4) extracting information to support decision-making; (5) addressing moral hazard and asymmetric information problems; and (6) facilitating the sale of purchase of goods and services through a payment system. There is much overlap in these descriptions. To setup a Collateralized-Debt-Obligation (CDO) one pools funds, manages risks, and moves funds across time. The BIS scheme identifies the functions performed by innovation, focusing on the transfer of risks (both price and credit), the enhancement of liquidity, and the generation of funds to support enterprizes (through credit and equity.). No commonly accepted and unique taxonomy of functions has been adopted. If functions represent timeless demands put upon financial systems, then why do we observe innovation? Some authors adopt a static framework, where no attempt is made to explain the timing of the innovation. Other authors adopt a dynamic framework, where innovations reflect responses to changes in the environment, and the timing of the innovation mirrors this change.

1.6 Summary: Impact of Regulatory Changes on Banking

1.7 Global Risk Capacity

1.8 Pricing, Hedging

Pricing and hedging of financial products are key competencies in financial innovation. The importance of these skills distinguishes the financial sector from other industry sectors. Pricing and hedging is much more demanding in the financial industry due to the temporal or forward looking properties of financial contracts. Issuing such contracts for the clients requires for the issuer to master uncertain or risky cash flows. Not all financial services are forward looking. The pricing of pure adminstration services or advisory services (financial planning, pension planning) are not considered.

There are three approaches to pricing.

- The integrated **economic approach** using a fully fledged economic model. Solving the model, i.e. finding the optimal policies for all agents in the economy under several constraints such as individual budget constraints and the market clearing conditions delivers the optimal individual decisions and from market clearing, the price dynamics of the asset in the economy are specified.

- The **no arbitrage approach**. This approach starts with the assumption that arbitrage is not possible: An investment strategy starting with zero initial wealth ending with certainty with no loss and in at least one state with a gain defines an arbitrage strategy. Such strategies are ruled out. Using this assumption and that people like more to less money relative pricing of assets follows: Derivatives are priced given the exogenous base assets such as stocks or bonds. This relative approach is neither contradictional nor orthogonal to the first general economic approach. In fact, no arbitrage is a necessary condition that financial equilibria exist. We discuss this in the next chapter.

- If prices of base assets do not exist since markets are in a state of infancy in practice a **potpourri** of different methods is used: Ad hoc rules, try-and-error approaches, signaling pricing and others apply.

We consider arbitrage pricing in some details in the next chapter. We therefore focus on general economic pricing. The difference between the first two approaches can be considered as follows. While economists populate the first approach, in the second one many physicists, mathematicians and other non-economists started to work in the last two decades ('Quants'). Banks were and are willing to pay high salaries to quants; higher ones than to comparable jobs which required an economic curriculum. The interdependence between these two groups is small. The research 'Econophysics' is an example that quants and economists often are not communicating. While quants pretend that their research is closer to reality since in their opinion they use less ambiguous economic concepts (utility functions), many economists argue that econophysics has little to do with economics but a lot with the application of physical concepts to economics.

Since ricing of financial products requires models, one faces **model risk** independent whether one considers the general equilibrium or the no arbitrage approach. Model risk sources can be possible misspecification of model parameters (utility functions, correlations), omissions of price sensitive variables, misspecification of estimates or uncertainty. Consider a new derivative product where no market exists for mark-to-market but mark-to-model is required for pricing. Traditional pricing models of Black and Scholes (equity), Black and LIBOR[43] market models (interest rates) or extensions of them apply. All these models face model risk, i.e. there are not enough payoffs to span the possible states of the world (market incompleteness), trading is restricted (short positioning is not possible) are two examples, volatility is a constant in the model but a function in reality, etc.

Pricing is only one side of the medal with **hedging** on the other side. Hedging applies to investment and trading product innovations where the issuer of the product faces a liability to the investor. The idea of hedging is: The initial price of a product is chosen such that the bank can invest in a portfolio which generates the payoff of the product in any possible states at any future dates. If this is possible, one speaks about a perfect hedge or **replication**. One assumes that the bank neither needs to inject additional money to cover the liability nor can it withdraw cash (self-financing). For other products such as bank deposits hedging is different. First, there are no tradeable products such that a bank deposit can be replicated. Second, risk management is defined on the whole balance sheet. That is interest -, liquidity - and credit risk are considered jointly for the asset and liability side. Hedging then means that the risk figures of the balance sheet are calculated and compared with the risk tolerance expressed by Value-at-Risk, Greek (key rate delta, convexity) and liquidity figures (NSFR) on an daily, operational basis

[43]London interbank offered rate (LIBOR) describes an interest rate which is published daily by the British Bankers Association. LIBOR is the average interest rate which banks in London are charging each other for borrowing. It's calculated by Thomson Reuters for the British Banking Association (BBA). It is a used benchmark for short term, i.e. up to 1 year, interest rates. LIBOR is offered in ten major currencies GBP, USD, EUR, JPY, CHF, CAD, AUD, DKK, SED, and NZD. LIBOR has been a factor in the pricing of hundreds of trillions of dollars of loans, securities and assets There are many vanilla LIBOR based instruments which are actively traded both on exchanges and over the counter such as LIBOR futures, forward rate agreements. The significance of these instruments is that: (a) They allow professionals effectively hedge their interest rates exposure. (b) One can use them to synthetically create desired future cash flows and thus effectively manage assets versus liabilities. (c) They allow market participants express their views on future levels of interest rates.

and with economic capital and risk budgeting on annual basis.

Hedging of investment and trading products requires that the innovator has a strong technology, strong human skills and risk capacity. Without a fast performing and secure IT innovation in financial products is no longer feasible. First, an overview over the positions is missing which means a blind flight in the management of risks follows. Second, without a performing infrastructure traders will lose money in particular if their is high flow due to market panics or market exaggerations. Third, derivative houses are under permanent attack of specialized software, so called high frequency trading. Say a traders set bid-ask offers at 100-101. Then the high frequency machine observes this and offers 100.01-100.99.

The pricing and hedging of derivative offer some intellectual challenges. We consider plain vanilla European call and put options on a liquid stock.[44] The pricing of the options follows from no arbitrage and mathematical reasoning in a perfect market: That is, there is a unique price formula for this options consistent with no arbitrage. Assuming zero dividend, the option price is driven by two parameters: The risk free interest rate and the volatility parameter. But it is not the historical volatility which matter - it is the **implied volatility** σ_{im}. By definition, this is the value which we put into the theoretical pricing formula to equalize observable market price:

$$\text{TheoPrice}(\sigma_{\text{im}}) = \text{MarketPrice} .$$

This requires an option pricing model such as the Black and Scholes model for example. Why is volatility a key parameter in trading? First, trading prices of vanilla options is the same than trading volatility - there is one-to-one relationship in the Black and Scholes model. An increasing volatility means increasing option prices and vice versa. We note that implied volatility for a call option by definition leads to the correct market price of the call but that if one inserts the same value in another option type, say a digital option, a wrong result follows, even if the options have the same underlying, the same maturity. As a second remark we observe that implied volatility is not constant. This parameter is a function of the maturity and the moneyness, i.e. how deep the option is in- or out-of-the-money. Implied volatility is indeed a surface in the two dimensions 'maturity' and 'moneyness'. Fixing one dimension, a volatility curve follows, see Figure 1.33.

Figure 1.33 shows the bid and ask volatility curves as a function of the moneyness. They show a typical shape for equity derivatives: A smile, i.e. U-shaped pattern, and a skew, i.e. a asymmetric smile. The figure shows that volatility increase with distance to the at-the-money (ATM) region, i.e. where actual stock price and strike price are close. This is typically the region of liquidity. The increase in volatility away from ATM indicates that uncertainty increases which make the option more expensive. Rebonato

[44] Plain vanilla means that the options do not possess a complicated structure such as trigger events, barriers or other path dependent features. European means that the option can only be exercised at maturity.

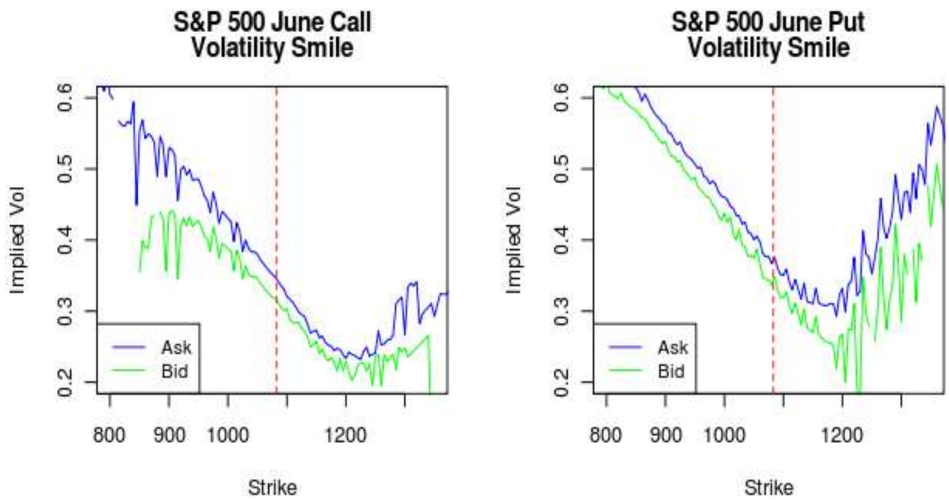


Figure 1.33: Volatility curves for fixed maturities for call and put options on Standard and Poor's 500 Index.

summarizes:

The implied volatility is the wrong number to put in the wrong formula to get the right price.

The traders use the calculated volatility curves from say Eurex options prices.

The Black and Scholes model (BSM) approach is widely used to price vanilla option. Even though volatility is not constant, the approach is consistent with market prices. But the model is no longer consistent for path-dependent options, i.e. options where the price not only depends on the probability distribution at expiry. For all path-dependent options BSM prices will not be in-line with observed market volatilities for relevant strikes and maturities. In order to bring prices for path-dependent options in-line with the market, option specific spreads are presently used on the vanilla option implied volatility surfaces.

Consider a barrier option. Then we face using BSM a problem which volatility we should plug into the analytical barrier option price formula - the implied volatility at the strike or barrier level in the volatility surface? One can choose one volatility and add/subtract a spread for the volatility at the other point. The problem of finding the single BSM volatility is exacerbated if the distance between the barrier and strike is large. The reason for the problem magnification is due to the curvature of the skew and often the large difference between strike and barrier volatility. If one consider other options, the problem accentuates, i.e. multiple barriers emerge, cash flows can enter, etc. The BSM then becomes unsuitable.

The Local Volatility Model (LVM) enable pricing of path-dependent option that is consistent with observed market volatility surfaces without the need of a volatility spread. The price is systematically determined by providing a set of volatility data instead of a single volatility value, as is the case with the BSM.

Local volatility models assume the following underlying price dynamics:

$$\mathrm{d}S(t) = rS(t)\,\mathrm{d}t + \sigma(S_t, t)S(t)\,\mathrm{d}W_S(t), \tag{1.10}$$

where σ is a **deterministic** function of both time and the underlying's price. This is opposite to stochastic volatility models where there is an own stochastic dynamics for the volatility state variable. Hence, the same risk source drives both the underlying and the volatility in LVM.

The motivation for this model class is as follows. For vanilla options, implied volatility $I(K,T) = I_{BS}$ is a function of strike and maturity. Hence, the fair vanilla option price is a function of K and T or in other words, $I(K,T)$ determines the risk neutral probability for vanilla option's underlying value $S(T)$ at maturity. The option value at maturity is independent of the underlying's path. If we consider a barrier option, not only the terminal value $S(T)$ matters for option pricing, i.e. pricing becomes path dependent.

The implied volatility surface for vanilla options will not tell the right answer, since the volatility figure is of a global type, i.e. path independent. The path dependency requires a volatility figure which is also path dependent, i.e. $\sigma(S_t, t)$ as for the local volatility model. Since this volatility is not a global figure but is only valid for given S_t and t, the expression "local volatility" is used.

Dupire (94) shows that, given a complete set of European option prices for all strikes and maturities, local volatilities are uniquely determined by the vanilla option prices and their derivatives. The Dupire equation describes the relationship between implied and local volatility. The non-discounted risk-neutral value $C = C(S_0, K, T)$ of a European call option is given by

$$C = \int_K^\infty \varphi(S_T, T)(S_T - K)\,\mathrm{d}S_T \tag{1.11}$$

where φ is the unknown probability density of the final spot price at maturity. Then,

$$\frac{\partial C}{\partial T} = \frac{\sigma^2 K^2}{2}\frac{\partial^2 C}{\partial K^2} - rK\frac{\partial C}{\partial K},$$

which is the Dupire equation with initial condition

$$C(K, 0) = (S(0) - K)^+ \ .$$

For the proof see Appendix 7.2. *Implied local volatility* $\sigma_{\mathrm{Loc}}(K, T)$ is then defined as:

$$\sigma^2_{\mathrm{Loc}}(K, T) = \frac{\frac{\partial C}{\partial T} + rK\frac{\partial C}{\partial K}}{\frac{K^2}{2}\frac{\partial^2 C}{\partial K^2}}. \tag{1.12}$$

This is the local volatility function consistent with the given prices of options and their sensitivities whereas the unknown density function ϕ has been eliminated. Equation (1.12) holds for non-dividend paying stocks. With a continuous dividend stream d, Dupire's equation reads:

$$\sigma^2_{\mathrm{Loc}}(K, T) = \frac{\frac{\partial C}{\partial T} + (r - d)K\frac{\partial C}{\partial K} + dC}{\frac{K^2}{2}\frac{\partial^2 C}{\partial K^2}}. \tag{1.13}$$

It follows that Dupire's equation is obtained by **switching** from the PDE in the variables S, t to a PDE in the variables K, T.
There is a simple interpretation of Dupire equation in terms of static option strategies. That for, we set the interest rate equal to zero. Then,

$$\sigma^2_{\mathrm{Loc}}(K, T) = \frac{\frac{\partial C}{\partial T}}{\frac{K^2}{2}\frac{\partial^2 C}{\partial K^2}} \ .$$

But $\frac{\partial C}{\partial T}$ is the infinitesimal version of

$$\frac{C(S, t, K, T + \Delta T) - C(S, t, K, T)}{\Delta T}$$

i.e. a long call position with maturity $T+\Delta T$ and a short call with maturity T. In other words, a calendar spread with strike K. Similar, $\frac{\partial^2 C}{\partial K^2}$ is the infinitesimal version of

$$\frac{C(S,t,K+\Delta K,T) - 2C(S,t,K,T) + C(S,t,K-\Delta K,T)}{(\Delta K)^2} .$$

But this is a butterfly spread with strike K. Hence, local variance is **proportional** to the ratio of a calendar and a butterfly spread.

Dupire equation looks much like the Black-Scholes equation with t replaced by T and S replaced by K. But whereas the Black-Scholes equation holds for any contingent claim on S, Dupire equation holds only for standard calls and puts. This equation tells you how to find $\sigma_{\text{Loc}}(K,T)$ and hence build an implied local volatility tree from options prices and their derivatives. You can then use that implied tree to value exotic options and to hedge standard options, knowing that you have one consistent model that values all standard options correctly rather than having to use several different inconsistent Black-Scholes models with different underlying volatilities.

Dupire's approach requires a continuous set of options data for all K and T. Since data are only available for a discrete set and options out- and in- the money are suffering from illiquidity several problems arise in the implementation of the approach. First, an interpolation between the discrete data points is needed. It turns out that different interpolation methods have a strong impact on the outcome. Besides this instability due to the interpolation, the calibration of the local volatility surface is also instable over time.

If we re-express Dupire equation in term of the original variable, by applying the chain rule and using the formula for the Greeks in Black and Scholes we get:

Proposition 1.8.1 (Local variance in terms of Black-Scholes implied variance)**.** *For zero dividends and zero interest rates, implied local variance reads in term of Black-Scholes implied variance:*

$$\sigma^2_{Loc}(K,T) = \frac{2\frac{\partial I_{BS}}{\partial T} + \frac{I_{BS}}{T-t}}{K^2\left(\frac{\partial^2 I_{BS}}{\partial K^2} - d_1\sqrt{T-t}\left(\frac{\partial I_{BS}}{\partial K}\right)^2 + \frac{1}{I_{BS}}\left(\frac{1}{K\sqrt{T-t}} + d_1\frac{\partial I_{BS}}{\partial K}\right)^2\right)} \tag{1.14}$$

with

$$d_1 = \frac{\ln(S/K) + (r + \frac{1}{2}I^2_{BS})(T-t)}{I_{BS}\sqrt{T-t}} .$$

Using the forward price and the transformation

$$x = \ln\left(\frac{K}{F_{0,T}}\right), \quad y(T,x) = I^2_{BS}(K,T|S,t)(T-t),$$

the result (1.14) reads:

$$\sigma^2_{Loc}(x,T) = \frac{\frac{\partial y}{\partial T}}{1 - \frac{x}{y}\frac{\partial y}{\partial x} + \frac{1}{4}\left(-\frac{1}{4} - \frac{1}{y} + \frac{x^2}{y^2}\right)\left(\frac{\partial y}{\partial x}\right)^2 + \frac{1}{2}\frac{\partial^2 y}{\partial x^2}}.$$

This proposition relates the two concepts: Implied local volatility and implied Black and Scholes volatility.

Proof. For further properties of LVM we refer to the literature, see Derman et al. (1994, 1996), Dupire (1996) and Gatheral (2006). There is in particular a relationship between local variance and instantaneous variance, i.e. the variance which is for example given in a stochastic volatility model such as Heston's one. We only assume that the underlying process is a diffusion - but no further specification is assumed:

$$\sigma^2_{\mathrm{Loc}} = \mathbb{E}\left[\sigma^2_T | S_T = K\right].$$

Implied local volatility is equal to the risk neutral expectation of the instantaneous variance conditional on the final stock price S_T being equal to the strike price K. In this way, (squared) local volatility can be thought of as the conditional, risk-neutral expectation of future variance. This is again analogous to forward rates which can be thought of as the risk-neutral expectation of future interest rates.

There are numerous advantages of using this model in practice. The most important of these is that it gives a description of volatility, without adding further sources of risk/Stochastic factor to the model. Therefore the model will still be complete. This is due to the LVM being a function only of the stock price and time. This leads to many nice features of the model. One of these features, and the second major advantage of the model, is that all European options can be fully hedged using only the underlying and risk less bonds (as was the case in the BSM). This means that many of the useful features of the BSM are preserved in the LVM. The third advantage of the model is that it is relatively simple since it only has one stochastic factor, making numerical implementation fairly easy. Finally, it is important to note that for a sufficiently smooth IVS, the model guarantees the existence of a unique LVM, meaning that knowledge of the IVS is equivalent to knowledge of the LVM. The model also has a few drawbacks. The first of these is that it gives wrong predictions of future volatility. This can be seen by comparing the model's prediction today of volatility at some future date, to the model's prediction of volatility at that future date after reconstructing the local volatility surface at that point. These two values will not necessarily be the same, meaning that the model's prediction of stock price dynamics can be inconsistent. The local volatility surface also tends to flatten out over time, meaning that the model predicts that the implied volatilities will flatten out over time. This is not empirically consistent and so the model's representation of stock price dynamics is not perfect. The second drawback of the model is that it predicts shifts in the IVS due to increases in the stock price that are contrary to typically observed movements in markets. More specifically, the model predicts skew shifts to the left for increases in the stock price, whereas in reality, shifts to the right are often observed. In spite of these drawbacks, the LVM can still be implemented effectively in practice, and it is possible to compensate for some of the problems with the model.

The LVM possesses two types of **model risk**.

- **Intrinsic model risk**: The LVM gives wrong predictions of future volatility. This can be seen by comparing the model's prediction today of volatility at some future date, to the model's prediction of volatility at that future date after reconstructing the local volatility surface at that point. These two values will not necessarily be the same. The local volatility surface also tends to flatten out over time, meaning that the model predicts that the implied volatilities will flatten out over time. This is not empirically consistent. Due to the bad predictive ability and the flattening of the surface the LVM falls short in volatility dynamics modelling. It is therefore not very suitable for options which have heavy dependency on the dynamics of the volatility, e.g. forward starting options.

- **Competitiveness model risk**: Although the LVM dominates other models such as BSM for path dependent options in terms of pricing and hedging accuracy, there is lower intrinsic model risk in LVM, total model while using LVM can be substantially larger if the competitors which are needed to buy and sell options do not use the LVM model. That is, if market practice uses prices which are different than LVM prices and rejects to trade at LVM prices.

Chapter 2

Discount Factors and No Arbitrage - The Basics

Pricing of assets is key in financial innovation. This brings us to the roots of finance: The time value of money or discount factors. We consider first the case of non-stochastic discount factors. We extend this to the case with risk and relate the stochastic discount factors to the no arbitrage paradigm.

2.1 Discount Factors - ad hoc View

Since one Swiss Franc today is in general worth more than 1 CHF in the future, cash flows at different time points cannot simply be added: There is a time value of money. Cash flows can be pushed forward or backward in time if their time value is considered. Discount factors precisely provide this information. Consider a future date T and present date $t \leq T$. A discount factor $D(t,T)$ states how much 1 CHF at future date T is worth at time t. If interest rates are positive, the discount factor satisfies $0 < D(t,T) \leq 1$. Given the discount factors we can value cash flow streams due to their additivity. The construction of discount factors is therefore a major task in pricing financial contracts.
Using discount factors allows one to reduce the complexity of cash flow profiles. The simplest way is to shift all cash flow to one particular date and to add the values. The present or net present value (PV or NPV) is such a measure which values all future cash flows discounted back to the present date. If all cash flows are projected onto a future date, the value is called the future value (FV). The PV and FV are used since they reduce a complicated multi dimensional problem - the comparison of whole cash flow streams - to a one dimensional comparison of numbers. This makes **pricing** of contracts possible. If we can construct the discount factors for different maturities T one can value and compare all cash flow profiles using the PV or FV.

There are different ways to express discount factors. This is largely due to different ways of interest rate compounding. Using continuous compounding the discount factor reads

$$D(t,T) = D(T-t) = e^{-(T-t)\times R}$$

with R the nominal interest rate. Taking the time derivative $\tau := T - t$ we get that the growth rate of the time value of money is the interest rate R. What kind of information does R reflect? So far, the discount factor captured the time value of money in a stylized world. But discount factors depend on different factors. Using compounding[1] they read as:

$$D(0,T) = \frac{1}{(1+R)^T} = \frac{1}{(1+R(\text{risk free, liquidity, credit risk}))^T} \ .$$

A standard way to approach the complicated and not known function R is to expand it using a Taylor series, i.e.

$$D(0,T) = \frac{1}{(1+R(\text{risk free, liquidity, credit risk}))^T} \sim \frac{1}{(1+R(\text{risk free }) + \frac{\partial R}{\partial L}\Delta L + \frac{\partial R}{\partial C}\Delta C)^T} \tag{2.1}$$

with L representing liquidity and C credit risk. This is a first order approximation where the terms which are added to the risk free rate are called the credit or liquidity spread. If the liquidity and credit risk interdependent one can consider the next order approximation to obtain this cross effects.

The existence of a risk free rate is a fiction. There are market circumstances where some rates can be considered to be risk free such as default risk for some government bonds. The near past however shows that what is assumed to be risk free today can be risky tomorrow. Even cash is not risk free - inflation and currency risk also affect the value of cash. Before the financial crisis the LIBOR rates were often considered to be quasi risk free - reflecting the believe that the large international active banks which fix LIBOR rates are default free. In the middle of 2012 furthermore the LIBOR fixing scandal popped up: More than a dozen banks are under investigation by authorities in Europe, Japan and the United States over the suspected rigging of the LIBOR in the period 2005 - 2009. calls the risk free assets to be money market accounts. In the financial turmoil of 2008 it became clear the money market accounts faced default risk of the respective banks - without the interventions of the governments which put liquidity into the banking system and gave implicit or explicit state guarantees there would have been considerable losses in such money market investments. In theory one likes to consider financial market models where there is one risk free asset and several risky ones. First, there can be no more than one risk free asset, else arbitrage holds. The assumption of a risk free asset is possibly driven by psychology and elegance. The existence of a risk free asset defines a benchmark both for modelling and for practitioners: The return of a strategy or a portfolio is then the risk free return plus 'a sum of risky returns'. That is

[1] Compounding means the way how cash flows are considered during the investment period. If they are consumed - simple compounding follows. If they are reinvested we call it compounding.

the risk free rate is the opposite to risky rates.

How many discount factors are there in a given currency for fixed dates t and T?

There exists only a **single** discount factor $D(t, T)$ for each economic environment.

If this is not true, we construct a money machine, i.e. **arbitrage** exists. Assume that there are two discount factors for the Euro in else two identical economic situations. One discount factor states that 1 Euro tomorrow equals 0.8 Euro today and the other one 0.9. Buy (a long position) the 0.8 zero bond and sell (a short position) the 0.9 one at t. Since both pay 1 tomorrow we make the risk less profit of 0.1 per unit Euro.

A second main implications of existing interest rates is on investment and loan business. Consider an agent with 100 savings in a currency an positive interest rates. If the agents do not invests the real value of its savings decreases over time. The amount of goods for consumption becomes smaller from year to year. Hence, there is an incentives to search for investment products which at least offset the decreasing purchasing power. Consider in the second case a bank which lends 100 to a borrower for a given period. If the bank does not charges interest rate payments the borrower will pay back in the future an amount which is in nominal terms still 100 but which in real units is lower.

Although discount factors are unique interest rates are not. The non-uniqueness of interest rates is related to how individuals behave in the different time scales which matter in the time value of money. The different time scales and human behavior are:

- **Maturity** of the contract T defines a time scale.
- **Basis of interest rates**, i.e. are the rates calculated on an annual, monthly or daily basis?
- Individuals can act in two different ways: Cumulated interest in a given period can be consumed or reinvested. The first behavior leads to **simple compounding** interest rate calculation, the second one to **compounding**.

In the calculation of interest payments one further distinguish between market rates (LIBOR), internal rates (Yield-to-maturity YtM) and day count conventions: How many days has one year and the how are the how do we count the days if accrued interest occur, i.e. how is "$dd.mm.yyyy - dd.m\tilde{m}.yyyy$" defined.

Interest rates are **always quoted per annum (p.a.)** and nominal interest rates are quotes. Simple compounding is used for structured products with maturity smaller than one year, for the LIBOR market interest rates and for accrued interest. Compounding is applied to saving accounts and bonds. Continuous compounding is used for derivatives and Treasury Bills for example. We note that the time value of money is always given by a difference between two dates $(T - t)$. **Time homogeneity** means that the future is not different than the present or that time itself matters only in a relative way in the time value of money. If one buys a financial product today the price remains unchanged if one buys the product at a future date if the economic conditions are unchanged. This assumption is conflicting with modern research about how people make consumption or investment decisions. This research indicates that decisions depend on the actual date or vista time: The discount factor is of the form $D(s, t, T)$ with s the vista-time of valuation. As an example consider a family spending their summer holidays at the sea side. The child is then asked to make the following choice: He can have one ice cream in 5 minutes or two of them in 1 hour. He also has to make the choice now whether in one year he would then prefer to wait 5 minutes for one ice cream or one hour for two of them. Impatience is likely to lead to the decision to prefer the immediate ice cream now and to prefer in one year two of them. Consider the family back in one year at the sea side under the same circumstances. The child is asked to make the same choice as in the past year. He then prefers only to wait 5 minutes and not one hour. He is conflicting his own decision made one year before. This is an example of an intra-personal conflict or time-inconsistent preferences. Such a decision problem cannot be modelled using traditional time homogeneous discount factors. There are situations were individual preferences are more presently oriented than the usual exponential discounting assume. The traditional form of discounting implies consistent time preferences, i.e. if an individual is asked to value two future events it does not matter when the individual is asked for the valuation given that the circumstances remain unchanged. In the case of more presently oriented preferences this is no longer true, see Rabin and O'Donoghue (1999). Individual are often aware of the future intra-personal conflicts and use different strategies to avoid them such as self- or pre-commitment strategies. Suppose that you know or would feel better if you would clean up your apartment. Although you plan to do this tomorrow, tomorrow is repeatedly postponed. To commit yourself you decide to invite for the next weekend guests which then breaks the inconsistency. An elegant formalism of such presently based preferences is due to Laibson (1997). Let u be the utility at a given date and U the utility over a period. We set

$$U^t(u_t, u_{t+1}, \ldots, u_T) = D^t u_t + \beta \sum_{s=t+1}^{T} D^s u_s$$

with $\beta > 1$ and D the discount factor. If $\beta = 1$, we are back in the exponential discounting case. If u depends on the consumption rate c for $\beta > 1$ we get from vista time t the following marginal utility for consumption at the future date $t+1$: $\beta D^{t+1} u'_{t+1}$. At the next day $t+1$ the optimal plan of the individual is $D^{t+1} u'_{t+1}$. This shows the conflict of the individual if for unchanged circumstances the only difference is the vista time of decision making. Vista time explicitly enters in discounting. Although this kind of research is interesting for individual decision making or even macro economic policy it is not used to price financial contracts. The main reason is that cash flow valuation depends on individual time preferences and the question then is: Whose time preference should matter? Second, suppose that some or even a majority of people fail to decide in a time consistent way. Is this a convincing argument that one should adopt this behavior and price say products in a non-time consistent way too?

Examples

1. We discuss the valuation of a single floating LIBOR cash flow C. C is fixed at date T_1 and payed at T_2. Calendar date is 0. The two dates T_i are distant exactly six months, i.e. we consider 6m LIBOR. We don't know at 0 what will be the value of 6m LIBOR in T_2. The best guess is the forward rate $F(0, T_1, T_2)$. We derive this rate using

$$\begin{aligned} \mathrm{PV}(C) &= D(0, T_1) \frac{C}{1 + F(0, T_1, T_2)/100 \times (T_2 - T_1)/360} \\ &= D(0, T_1) D_F(0, T_1, T_2) C \end{aligned} \tag{2.2}$$

i.e. the cash flow C is first discounted from T_2 to T_1 using the forward rate and then back to 0 using a spot rate discount factor. The term $(T_2 - T_1)/360$ represents the day count convention 'Actual/360'. The division by 100 implies that the forward is a percentage value. On the other hand, we have

$$PV(C) = D(0, T_2) C \ . \tag{2.3}$$

No arbitrage implies

$$D(0, T_1) D_F(0, T_1, T_2) = D(0, T_2)$$

. Solving for the forward rate:

$$F(0, T_1, T_2) = \left(\frac{D(0, T_1)}{D(0, T_2)} - 1 \right) \times \frac{36'000}{T_2 - T_1} = \left(\frac{D(0, T_1)}{D(0, T_2)} - 1 \right) \times \frac{1}{\alpha_{1,2}} \ , \tag{2.4}$$

with $\alpha_{1,2}$ the accrued interest factor.

2. We consider an investor which buys a capital protected structure product from a bank (the issuer). The investments is 100 CHF and the product matures after 5 years. The payoff of the product consists of a capital protection part which pays 100 CHF after five years conditional that the bank is not bankrupt and a participation at a given stock. To guarantee capital protection, the bank invests the PV of 100 buying a zero coupon bond, i.e. an instrument which pays 100 in 5 years and zero coupons in between. The bank has two alternatives to buy the zero bond from two other banks with different creditworthiness. bank 1 has a better rating than bank 2. Therefore, bank 1 pays interest rates of one-year Swap plus 2%. Bank 2 instead pays one-year Swap plus 3%. We assume that one-year Swap is 1%. The inverse 5-years discounting factors for the two banks using simple compounding are

$$D_1^{-1} = 1 + 5(1\% + 2\%) = 1.15 \ , \ D_2^{-1} = 1 + 5(1\% + 3\%) = 1.2 \ .$$

Therefore, the PV for the capital protection part is:

$$PV_1(100) = \frac{100}{1.15} \ , \ PV_2(100) = \frac{100}{1.2} \ .$$

If the structured product issuer does not account for the different creditworthiness of the zero bond issuers, buying the bond from bank 2 allows him to show a higher participation rate for the investor since the difference between the investment of 100 and the PV is used to buy derivatives which generate the participation at the chosen underlying value. In this case the investor an unhedged higher credit risk exposure. To see this, let $P(B|Y_i)$ be the probability that the issuer B defaults given that bank $i, i = 1, 2$ defaults. Assuming $P(B|Y_1) = P(B|Y_2)$, $P(B \cap Y_1) = 1/100, P(Y_1) = 2/100$ and $P(B_2) = 5/100$ it follows from the equivalence

$$P(B|Y_2) = \frac{1/100 \times 5/100}{2/100} = 5/200 \ .$$

Therefore, if the issuer funds the structured product from bank 1 we get the default probability

$$P_1(B) = 3/100^2 \ ,$$

and funding from bank 2 gives

$$P_2(B) = 12.5/100^2 \ .$$

The probability of default of the issuer is about 4 times larger if the higher participation product is issued compared to the other case. The investor faces a four time larger risk that the promised capital protection evaporates. Hence, to account for this difference in funding, the discount factors should contain a premium for credit risk. The situation can become severe if the issuer pushes business with the bank 2 without considering counter party risk in the pricing of the structured products. In this case, the relation $P(B|Y_1) = P(B|Y_2)$ is changed into $L \times P(B|Y_1) = P(B|Y_2)$ with $L > 1$. That is the issuer has a leveraged position. The leverage multiplies through the value chain, i.e. it multiplies the credit risk of the final investor in the structured product.

3. So far we neglected **bid/ask spreads**. Consider the no arbitrage relation between forward rates and zero bonds: $1 + F(0, T_1, T_2) = \frac{p(0,T_1)}{p(0,T_2)}$. Replicating the forward rate means buying the T_1 bond and selling short T_2 bonds. Therefore, the ask price $F^a(0, T_1, T_2)$ uses the ask price of the T_1 bond and the bid price of the T_2 bond, i.e.

$$1 + F^a(0, T_1, T_2) = \frac{p^a(0, T_1)}{p^b(0, T_2)} \ .$$

Similarly the bid price of the forward rate satisfies

$$1 + F^b(0, T_1, T_2) = \frac{p^b(0, T_1)}{p^a(0, T_2)} \ .$$

$F^b < F^a$ implies consistency conditions under no arbitrage:

$$p^b(0,T_1)p^b(0,T_2) < p^a(0,T_1)p^a(0,T_2) \ .$$

4. **Negative interest rates**. In some currencies such as CHF one observes 2011 and 2012 negative interest rate LIBOR fixings for 1m or even 3m maturity. This means that 1 CHF is worth less today than say in 3m. Suppose that you lent in CHF. Does this means that the bank has to pay you for the money lent to you? Although this is correct from an economic point of view it is not from a legal one since a negative interest rate violates the meaning of a loan - the borrower has to compensate the lender and not vice versa. A difficulty arises if documentations are not properly formulated. Consider an investment product with capital guarantee and a participation at the LIBOR rate. If the rate is negative, capital is no longer guaranteed. Hence banks which wrote LIBOR in their termsheets and not $\max(\text{LIBOR}, 0)$ are likely to face claims of investors if after maturity the capital is not repayed at the guaranteed level. A third problem are interest rate models which are used for markt-to-model pricing of interest rate derivatives such as caps and floors. The market standard for these vanilla options is the Black model. This is a model of the so called log normal type. Basically, the logarithm is not defined for negative arguments on the real axis. That is, negative rates lead to a break down of theoretical option pricing. A final problem follows if the LIBOR rate is used to create synthetic positions where in case of negative rates different procedures apply to the LIBOR rate. Consider a corporate client which asks for a fixed mortgage loan. This demand is replicated synthetically as follows:

   - The client enters in a 3m LIBOR roll over loan. They pay 3m LIBOR which is settled each 3m plus a fee.
   - The client also enters in a Swap contract with the bank: He pays a fixed rate and receives the 3m LIBOR rate.
   - Netting these two contracts implies that the client pays the fixed swap rate and the fee, i.e. the LIBOR components cancel.

   The problem with negative rates shows up as follows. The loan transaction in 3m LIBOR and the payer Swap transaction with 3m LIBOR floating leg are in two different markets. The first one is in the internal corporate banking market, the second one in trading department which acts on the capital markets. Therefore, if say in the first markets by the legal view LIBOR cannot be negative for a loan this view is not applicable in the capital markets: The two LIBOR positions no longer cancel. A profit and loss follows. The questions is then who is the owner of the profit and loss - the bank or the client?

5. **Inverse term structure.** Figure 2.1 shows the price of credit risk for ABB company in the period Aug 03 until Aug 04. The price of credit risk is expressed for Credit Default Swaps (CDS) derivatives in bps for different maturities (term structure). The price reflects how much a protection buyer has to pay to a protection

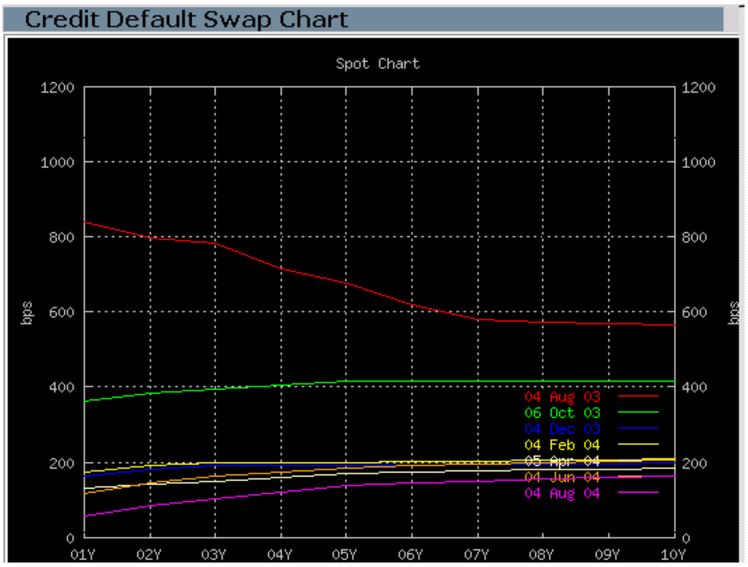


Figure 2.1: Term Structure dynamics for ABB CDS risk. On the horizontal axis are the maturities for the CDS contracts and on the vertical axis the prices in bps p.a. *Source: JP Morgan ORBIT.*

seller if the buyer wishes protection in case of ABB default for a given maturity. The top line shows that in Aug 03 the term structure was inverse. This reflects that in the CDS market actors believed that if ABB is surviving the nearest future then it will also survive later on. That is, a situation comparable to an intensive care case where the probability to die are higher for the short run than for the long run. ABB faced in this period serious problems: They were hit by general economic uncertainty, they faced asbestos claims and several mismanagement decisions had an impact. The stock price was down from over CHF 40 to less than 1 Swiss Franc. Since ABB could handle the problems, the price of credit risk became normal for latter term structures.

2.2 Stochastic Discount Factor - Asset Pricing

Stochastic discount factors (SDF) generalize the discount factors of last section, i.e. the discount factors of last section are contained as a special case in the more general setup. We follow Cochrane (2005) and Telmer (2007). Besides the introduction of risk we also derive the SDF from first economic principles leaving the former ad hoc discussion. In analogy to the short introduction to presently oriented preferences the time value of money follows from human behavior using a micro economic specification. Let u be the utility derived from consumption of a good c which is a smooth function, e the endowement S the price of a an asset. We consider decision making in a discrete time setup. Since preferences are time-separable it is sufficient to consider the decision problem for the consumer at date t and $t+1$ - the general problem beeing a evident generalization. The consumer chooses the portfolio ϕ_t, i.e. the amount of asset, such that expected utility is maximized and the budget constraints hold true:

$$\begin{aligned} \max_{\phi_t} & \quad u(c_t) + E_t[bu(c_{t+1})] \\ c_t &= e_t - \phi_t S_t \\ c_{t+1} &= e_{t+1} + \phi_t S_{t+1} \ . \end{aligned} \tag{2.5}$$

$b > 0$ describes the subjective impatience of the investor and the expectation is conditional on the information available at time t. The randomness enters from the stochastic prices of the asset. The budget constraints are binding since there is no opportunity than consumption. We only solve for the investment strategy. A full analysis would also consider the optimal choice of consumption. We do not carry this out, i.e. the optimal consumption path will enter implicitly the results. Inserting the budget constraints in the objective function and deriving the first order conditions w.r.t. ϕ implies the optimality condition (FOC)[2]:

$$S_t u'_t = E_t[bu'_{t+1} S_{t+1}] \text{ ,or } S_t = E_t[\frac{bu'_{t+1}}{u'_t} S_{t+1}] \tag{2.6}$$

where the prime denotes the derivative w.r.t. consumption. The condition expresses the marginal condition for an optimum. At time t there is a loss in utility if the investor buys another unit of the asset. This loss is equalized to the expected gain in the next period due to the additional payoff. If we define the SDF

$$D(t, t+1) := \frac{bu'_{t+1}}{u'_t} \ , \tag{2.7}$$

the **basic asset pricing equation** reads

$$S_t = E_t[D(t, t+1) S_{t+1}] \ . \tag{2.8}$$

Since $D(t, t) = 1$, the discounted price process is a martingale. This implies that it has no drift or the expectation is constant. This is due to the simplicity of the optimization problem, i.e. absence of a dynamics of the asset prices for example. If there is no risk, the asset pricing formula becomes the standard NPV formula. For discrete comounding we get $D(t, t+1) = \frac{1}{1+r} = \frac{bu'_{t+1}}{u'_t}$. Therefore

$$(1+r)\frac{bu'_{t+1}}{u'_t} = 1 \ .$$

[2]The FOC follow from:

$$\frac{\partial}{\partial \phi_t} \left(u(e_t - \phi_t S_t) + E_t[bu(e_{t+1} + \phi_t S_{t+1})\right) = 0 \ .$$

This shows that time impatience and the substition rate are neutralized by the interest rate. The pricing formula tells us that there is a correlation between the random components of the common discount factor and the asset-specific risk. The SDF is the marginal rate of substitution, i.e. the rate at which the investor is willing to substitute consumption at future time for consumption at a present time. The pricing formula can refer to real (denominated in goods) or nominal (denominated in dollars) prices. The difference is only in the SDF being real or nominal. If prices are nominal values, real prices follow by using a numeraire, here a price index N_t, i.e.

$$\frac{S_t}{N_t} = E_t[D(t,t+1)\frac{S_{t+1}}{N_{t+1}}] . \tag{2.9}$$

The normalization can be packed into the nominal disount factor D leading to a real discount factor D'. The basic return equation follows at once by taking S_t inside of the expectation in (2.8):

$$1 = E_t[D(t,t+1)(1+r_{t+1})] . \tag{2.10}$$

Consider several assets. For each asset the return equation (2.10) holds. Pick assets j and k. We get

$$0 = E_t[D(t,t+1)(r^j_{t+1} - r^k_{t+1})] . \tag{2.11}$$

Since the expectation is a scalar product, i.e. $E(fg) = \langle f, g\rangle$, the excess return are orthogonal to the SDF. This is in particular true if one asset is the risk free one. This leads us directly to risk premia, i.e. the conditional risk premium of asset k is the expected excess return over the risk-free rate r_t: $E_t[r^k_{t+1} - r_t]$. The unconditional risk premium follows by taking expectation of the conditional one and using the tower property, i.e. we get $E[r^k_{t+1} - r_t]$. Since $E(fg) = E(f)E(g) + \mathrm{Cov}(f,g)$, we get from (2.11)

$$E_t[D(t,t+1)]E_t[(r^j_{t+1} - r_t)] = -\mathrm{Cov}_t(D(t,t+1), r^j_{t+1}) . \tag{2.12}$$

Assets with returns which covary positively with the SDF will pay a negative risk premium. If the SDF is constant, covariance is zero and risk premia (unconditional and contional one) are zero. As a first application the law of iterated expectations (tower law) implies

$$E[(r^j_{t+1} - r_t)] = -\frac{\mathrm{Cov}(D(t,t+1), r^j_{t+1})}{E[D(t,t+1)]} . \tag{2.13}$$

We drop in this case the time indices. If we rewrite the above condition for a benchmark portoflio return r^* and divide the two expressions we get

$$\frac{E(r^j - r)}{E(r^* - r)} = \frac{\mathrm{cov}(D, r^j)}{\mathrm{cov}(D, r^*)} =: \beta^j ,$$

i.e.

$$E(r^j - r) = \beta^j E(r^* - r) . \tag{2.14}$$

If we further specifiy $D = a + br^*$ for the discount factor $\beta^j = \frac{\text{cov}(r^*,r^j}{\text{var}(r^*)}$ follows. For r^* the market portfolio, the above specification is the CAPM. As a second application we consider again the case where the covariance between the SDF and the asset is zero. Then no matter how large the idiosyncratic asset risk is, its volatility, it has no impact on the asset pricing. Only systematic risk generates a risk correction. We finally apply the result to mean-variance analysis. If R is a gross return of an asset, we get from the fundamental asset pricing equation

$$1 = E[DR] = E[D]E[R] + \rho_{D,R}\sigma(R)\sigma(D) \ . \tag{2.15}$$

This implies

$$E[R] = \frac{1}{E[D]} - \rho_{D,R}\sigma(R)\frac{\sigma(D)}{E[D]} \ .$$

Since $\frac{1}{E[D]}$ equals the risk free rate r and the absolute value of the correlation is bounded by 1 we get the bound

$$|E[R] - r| \leq \sigma(R)\frac{\sigma(D)}{E[D]} \ . \tag{2.16}$$

Hence, the set of means and variances of returns is limited. They must lie in the wedge-shaped region illustrated in Figure 2.2. The boundary is the **mean-variance frontier**, i.e. the region which tells the investor 'how much mean return he can get for a given level of risk'. The frontiers correspond to the unit absolute value correlation. Returns on the upper part of the frontier are perfectly negatively correlated with the discount factor and hence positively correlated with consumption. Theyare 'maximally risky' and thus get the highest expected returns. Returns on the lower part of the frontier are perfectly positively correlated with the discount factor and hence negatively correlated with consumption. They thus provide the best insurance against consumption fluctuations. All frontier returns are also perfectly correlated with each other, since they are all perfectly correlated with the discount factor. This fact implies that we can span or synthesize anyfrontier return from two such returns. Since each point on the mean-variance frontier is perfectly correlated with the discount factor, we must be able to pick constants a, b, c, d such that $D = a + bR^{mv}, R^{mv} = d + eD$. Thus, any mean-variance efficient return carries all pricing information. Given a mean-variance efficient return and the risk-free rate, we can find a discount factor that prices all assets and vice versa. Thus, any mean-variance efficient return carries all pricing information. Given a mean-variance efficient return and the risk-free rate, we can find a discount factor that prices all assets and vice versa.

2.3 No Arbitrage

The notion of no arbitrage enters in the pricing of assets, in particular in the pricing of derivatives such as swaps, structured products, real estate derivatives, credit default swaps, warrants. Pricing of assets is roughly driven by the no arbitrage principle.

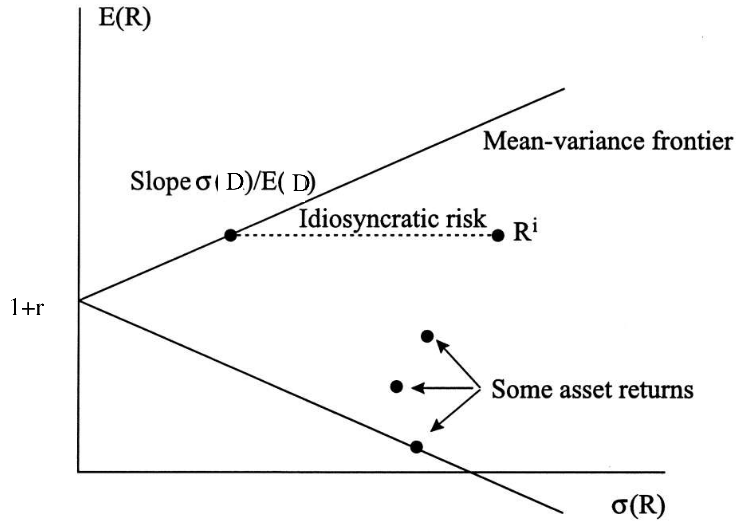


Figure 2.2: Mean-variance frontier. The mean and standard deviation of all assets priced by a discount factor m must lie in the wedge-shaped region. *Source: Cochrane (2005).*

The strength of this approach is that a unique principle leads by logic reasoning to the prices of derivatives. The theory of no arbitrage pricing has obtained indeed a mathematical foundation culminating in the work of Delbaen and Schachermayer (1994). Fortunately, the main ideas can be explained using simple mathematics. We consider the whole section the simplest possible setup - **one period, discrete states**.

We consider perfect markets, i.e. we assume:

- There are no transaction costs or taxes.
- All securities are perfectly divisible (i.e. it is possible to buy any fraction of a share).
- There are no restrictions on short selling.
- Markets are liquid and there is no liquidity premium.

There are three approaches to the pricing of derivatives which are all based on the no arbitrage principle:

- Replication approach.
- Hedging approach.

- Risk neutral pricing.

A different, more absolute approach is the **(general) equilibrium** approach. The no arbitrage approach prices assets based on prices of other assets - derivative are priced given the prices of the underlying values and the use of the no arbitrage principle. The no arbitrage approach uses minimal behavioral assumptions. The approach cannot detect mispriced underlying assets neither can one understand the implications changes in taxation or other institutional characteristic on asset prices.

The equilibrium approach is appropriate when we want to value the primitive securities. To do this we must understand what ultimately drives the supply and demand for the risky cash flows: Strong assumptions about investor preferences, trading frictions, information structure are necessary. These strong assumptions make it difficult for equilibrium models to be successful in practice, i.e. where it is important to know a robust price of a derivative with precision.

The arbitrage approach is **consistent** with the equilibrium approach in that all properties of the former will hold in the latter. The converse, however, is not true.

To motivate the arbitrage approach, we consider a **minimal market** with two dates, 0 and T.

- There is a stock S with price 100 in a given currency at time 0, see Figure 2.3. Research estimates that
  - the stock raises to a value of 120 with a probability of 90% and
  - that the stock drops with a probability of 10% to a price of 80.
- The investor can buy the following contract - a derivative C - at time 0 with the payoffs at time T
  - 20, if the stock rises,
  - 0, if the stock drops.

  This is a call option with strike 100 for the underlying S.
- There is a risk-less instrument B with price 1 today and which pays 1.1 at time T independent whether the stock rises or falls.

How **much** is the investor willing to **pay** at time 0 for the derivative C - this is the pricing problem.

We show that there is a **unique answer** to this question, which can be considered to be fair. To achieve this goal we introduce the motivations of a seller (writer) and of a buyer of the derivative.

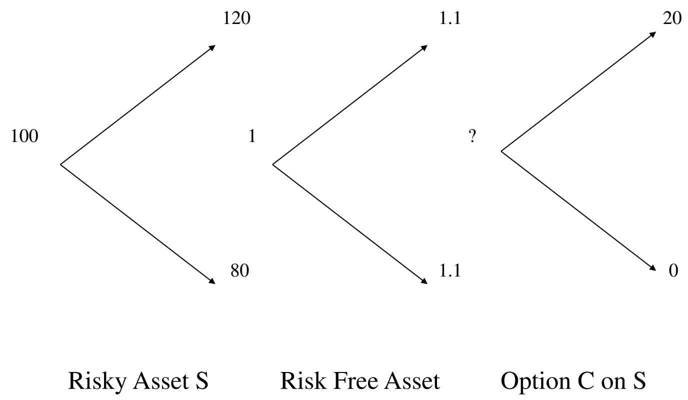


Figure 2.3: One period model.

The writer of the derivative would like to obtain a price from the buyer at time 0 such that he can buy a portfolio V_0 at 0 which will have a value V_T at time T which is always worth at least the value of the derivative C_T at time T, i.e.

$$V_T(\omega) \geq C_T(\omega) \quad , \quad \text{in all states } \omega \ .$$

The price at time 0 should be high enough that the writer can pay the liability at time T using the price change of the portfolio V_0 up to time V_T without additional money[3] and using the three instruments S, B, C only. The buyer of the derivative wants not to pay a price at 0 for the derivative such that the writer can buy a portfolio V at time 0 which is worth more than the derivative value at time T, i.e.

$$V_T(\omega) \leq C_T(\omega) \quad , \quad \text{in all states } \omega \ ,$$

is the buyer's intention. The buyer does not want to pay too much for the derivative. The price, if it exists, where both motivations are met

$$V_T = C_T \quad , \quad \text{in all states } \omega \ .$$

is called the **fair price of the replication portfolio** (we often skip the state variable ω in the notation). There are no restrictions on the portfolio positions, i.e. we can be long or short any instrument in the market.

[3]The portfolio is required to be self-financing: All changes in the portfolio value from time 0 to time T are due to changes in asset price values in that period.

What is a 'state?' It represents everything that is relevant for the value of the firm, including firm-specific variables such as earnings and leverage, industry-specific variables such as product demand and input prices, and macroeconomic variables such as interest rates and exchange rates. The state includes everything that we are not going to model explicitly. Sometimes it includes the stock's price, so that, even though we are ultimately interested in deriving the stock price as a function of more primitive variables, the distinction between the state and the price becomes blurred.

Replication is not always possible. This leads to the following definitions.

- **Complete Market**: $V_T = C_T$ **holds in all states**. This is the case of replication or a perfect hedge.
- **Incomplete Markets**: $V_T = C_T$ **not always holds true**. The portfolio value at time T can be smaller or larger than the liability value in some states. A portfolio is called a hedge in such a setup and there exists always hedging risk.

The use of language is less precise in practice. There one always speaks about hedging, independent whether there is hedging risk. To find the replication portfolio V we buy or sell an amount A of the risky asset and buy or sell B risk less products. The condition $V_T = C_T$ is equivalent to two linear equations for A and B:

$$\begin{aligned} A * 120 + B * 1.1 &= 20 \\ A * 80 + B * 1.1 &= 0 \ . \end{aligned} \tag{2.17}$$

Solving the system we get

- $A = 0.5$, i.e. buy 1/2 risky asset.
- $B = -36.36$, i.e. long a loan with value 36.36.

Choosing A, B in this way, there are no hedge risks. What is the fair price $C(0)$? To answer this we calculate the portfolio value at time 0 using the above (A, B)-strategy:

$$V_0 = 0.5 * S_0 - 36.36 = 0.5 * 100 - 36.36 = 13.64 \ . \tag{2.18}$$

Is this also the fair derivative price? Yes. Indeed we apply the Law of One Price which is a weaker formulation than the no arbitrage principle:

Definition 2.3.1 (Law of One Price)**.** *Two assets with identical cash flows must trade at the same price or if the replication price of an option exists, then this price is unique.*

One often states the law of one price as follows: If we have three payoffs x, y, z at a given date with

$$z = x + y \ .$$

Then the prices $p(\cdot)$ at **any date** of these equal payoffs also agree:

$$p(z) = p(x) + p(y) \ .$$

If this is not true, one constructs money machines. Although this looks like a trivial idea its consequences are not - they define the innovations called **mathematical finance**, **financial engineering** and the concept is also key in **general equilibrium financial economics**.

The probabilities $P = (90\%, 10\%)$ do not matter in the pricing of the derivative. The fair price is - if the market is complete - independent on individual belief's of the market participants. This is a major reason for the success of derivative pricing. Given this observation people are tempted to state that the belief is of no importance at all. This is not true. Suppose that the common belief is that Google's stock price will raise by 10% in one week. Then the belief does not enter a derivative contract of Google but it clearly affects the level of the stock. Therefore, beliefs matter in derivative pricing by affecting the underlying's price level.

It is interesting to setup the above market as a game with other people and to ask them what is the price they would pay for the derivative. Given the price list, for each price different from the fair one, the arbitrage strategy is defined and the risk less profit is calculated. Two things are interesting in doing this. First, the list of prices is arbitrary. That is, most people have no idea about a method to determine the fair price. Second, the astonishment if they see how the writer of the option makes the risk less profits in a risky environment - the magic of option pricing plays at this stage.

Since $V_T = C_T$ we must have $V_0 = C_0$. If a different price follows, the writer can make risk less profits in a risky environment. For $V_0 < C_0$, the writer invests the difference in the risk less asset. If $C_0 < V_0$ the writer buys the derivative from the investor and sells it to the fair price. Again the difference is looked in and invested in the risk less asset. The law of one price is the most important special cases of no arbitrage.

Example

An investor builds up a stock position. He wants to possess 100 stocks in one year of a particular firm with today's stock price $S_0 = 100$. Assuming that the firm pays no dividend, he considers two investment opportunities:

- Opportunity A: He buys the stock on a forward basis at a price X. This contract needs no cash at time 0.
- Opportunity B: He buys the stock today. That for, he needs a loan of $10'000$ CHF with an interest rate of 5%.

In both opportunities, the value of the portfolio at time 0 is equal to zero. In the forward case, the value V_T is equal to the difference between spot and forward price X, i.e.

$$V_T = S_T - X \ .$$

The value V_T for the opportunity B is spot minus repayment of the loan, i.e.

$$V_T = S_T - S_0(1 + 5\%) \ .$$

Since the value of the two portfolios is the same at time 0, they have to be equal also at time T, i.e.

$$S_T - X = S_T - S_0(1 + 5\%) \ ,$$

i.e. $X = S_0(1+5\%) = 105$. In this case no arbitrage determines the price of the **forward** which is equal to the spot price plus the **cost-of-carry**. We apply this in the **Interest Rate Parity**. The interest parity is a basic equation which relates interest rates and FX rates. The parity is based on the no arbitrage principle. Although theoretical in nature the parity has wide applications in practice too. There are two parity relations.

- *Covered Parity*: The return of a domestic risk free investment equals the return of a foreign risk free investment if the FX risk is hedged using a forward contract.
- *Uncovered Parity*: The interest differential between two countries is compensated by the expected FX changes.

An arbitrage strategy tries to make money based on the uncovered parity. Such trades are called 'carry trades'. We consider first the covered parity and discuss it using Japanese yen (JPY) and Brazilian Real (BRL). If JPY are exchanged against BRL there is no guarantee that BRL does not devaluates. Using a FX forward we eliminate this risk. We assume

- Interest rate Yen $R_{JPY} = 1\%$ p.a.
- Interest rate Real $R_{BRL} = 10\%$ p.a.
- Spot Rate $S(t) = 0.025$ BRLJPY.

We consider two dates 0 and $T = 01y$ from a Japanese investor's view. The investor acts as follow in 0:

- He borrows JPY 1000 at 1% for 1y, i.e. he pays back JPY 1010 at T.
- He changes the JPY into BRL at the spot rate which gives BRL 25.
- He invests the BRL 25 at 10% for 1y, i.e. he receives at T BRL 27.50.

In $T = 1y$, the investor changes the BRL 27.50 into JPY at the spot $S(T)$ which is not known at 0: The above strategy is risky. To choose a risk free FX strategy he replaces the today unknown spot rate $S(T)$ by the known forward price $F(0,T)$. The forward price is determined with the above no arbitrage argument. This leads to the covered interest rate parity. We write R_d for the nominal interest rate in the domestic currency JPY and R_f for the interest rate in the foreign currency BRL for one year. Figure 2.4 illustrates the strategy where borrowing is in the foreign currency. At 0:

- The investor borrows BRL for one year. He exchanges the BRL at the spot $S(0)$ in JPY and invests the JPY for $1y$.
- He buys a forward $F(0,T)$ to exchange in one year JPY against BRL.

At T:

- The investor exchanges $(1+R_d) \times F(0,T)$ JPY in BRL.
- He pays back the borrowed BRL amount and pays $(1+R_f)\times$ BRL.

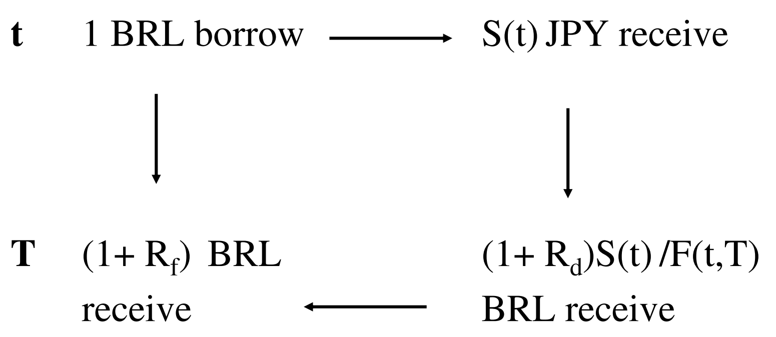


Figure 2.4: Representation of the forward strategy to hedge the FX risk.

To avoid arbitrage at time T the amount received in foreign currency cannot be larger than the amount of foreign currency payed back. Else a risk less profit follows. This implies the Covered Interest Rate Parity Theorem (CIP)

$$F(t,T) = S(t)\frac{(1+R_d)}{(1+R_f)} \,. \tag{2.19}$$

The difference

$$R_d - R_f$$

is the **interest rate differential**. The covered parity states that the difference between domestic and foreign interest rate determines the forward price. For lower domestic rates the forward price of the foreign currency is lower than the spot price.

Assuming $S(t) = 106$ JPYUSD, R_d (in JPY) $= 0.034$ and R_f (in USD) $= 0.050$. We get for $F(t, 1y)$

$$F(t, 1y) = 106 \times (1 - 0.016) \text{ JPY/USD} = 104.304 \text{ JPY/USD}.$$

If this price is violated, say a Bank offers the forward for 100 JPYUSD, then a U.S. investor exploits this by borrowing in the cheaper USD currency, changing this amount into Japanese Yen earning the foreign interest rate on this amount and he buys USD on a forward contract basis. At maturity he changes the Yen into USD at the forward price and pays back the loan.

A test of the CIP for CHFGBP are corrected delivers for the payoff

$$F(t,T) - S(t)\frac{(1+Rd)}{(1+Rf)}$$

a median value of -0.00099 for GBP-CHF and -0.00103 for CHF-GBP. This shows that the deviations are smaller than the bid ask spread. This indicates that it is difficult to develop no arbitrage strategies in the FX markets.

What is the difference between the uncovered and the covered parity? The difference is to replace the forward price in the above procedure leading to the CIP by the expected spot price, i.e. replace $F(t,T)$ by $E_t[S(T)]$ where this the conditional expectation given the information up to time t.

This leads to uncovered interest rate parity (UIP):

$$E_t[S(T)] = S(t)\frac{(1+R_d)}{(1+R_f)} \,. \tag{2.20}$$

Which view enters the expectation? If one beliefs that the best guess is the same as the forward rate, we are back to the CIP: There is no FX risk left. Carry trades are a bet that the expectations formation is different from the forward rate:

$$E_t[S(T)] \neq F(t,T) = S(t)\frac{(1+R_d)}{(1+R_f)} \ . \tag{2.21}$$

Consider a Swiss investor which needs in 30d JPY. As a first strategy, he buys the 30d JPYCHF forward instead buying the Yen at the spot. This fixes the exchange rate and he invests the money for 30d in CHF. This is a covered position, i.e. there is no FX risk. The higher interest rate in CHF is compensated by a discount in the forward. A second, different strategy is to exchange the CHF in JPY at the spot rate $S(t)$, to invest the amount in the Japanese money market for 30d and to pay the debt in JPY back. This also leads to a CIP. The lower Japanese interest rate is compensated by the difference between spot and forward price. Finally, a third strategy is to invest the CHF amount and to exchange it in 30d into JPY. This investment is not covered, i.e. FX risk is only zero if realized 30 spot equals the forward price.
This shows that if the forward is lower than indicated by the CIP one borrows money in the foreign currency, exchange it in domestic currency at the spot price and lend in the domestic currency. Contrary, if the forward is higher than CIP indicates, borrow in the domestic currency, exchange at the spot rate into domestic currency and lend in the foreign currency.

A currency carry trade is by definition a strategy to borrow in a currency with low interest rates and to invest simultaneously in a currency with high interest rates. This can only be profitable if the expected spot price and the forward price deviate, else the interest rate difference is compensated by the forward price difference. For JPY this means to borrow at close to zero Japanese interest rates and to invest in a currency with high rate. Assume that he Japanese rate is 0.5%. The loan in JPY is exchanged at spot prices into USD where USD interest rates for one year are 5.25%. If the exchange rate between USD-JPY remains unchanged, the net gain is 4.75%. If USD weakens relative to JPY the gain shrinks since one then needs more USD to repay the debt in JPY.

Does the UIP holds or equivalently, do carry trades make sense? We tested these questions and present the results below.

Figure 2.5 shows the distribution of the monthly returns for JPY-USD carry trades. The results show that in a majority of the cases the returns are positive. If UIP would hold true the return should be zero. The return are small, i.e. high capital amounts are needed to make substantial gains (or losses) in carry trades. The variation of the positive returns is smaller than for the negative ones. Hence, under normal market conditions the carry trade produces a small positive return. An analysis shows that the median return is 0.005 and that the distribution is skewed to the left: Therefore, with small probabilities high losses compared to gains are possible in a carry trade.

In the replication approach a portfolio of bonds and stocks was set up to replicate the derivative payoff. In the **hedging approach** one considers an unknown amount of the stock and the derivative. This portfolio is then specified by requiring that the portfolio has the same value in all states of the world as the risk less bond. Therefore, using the

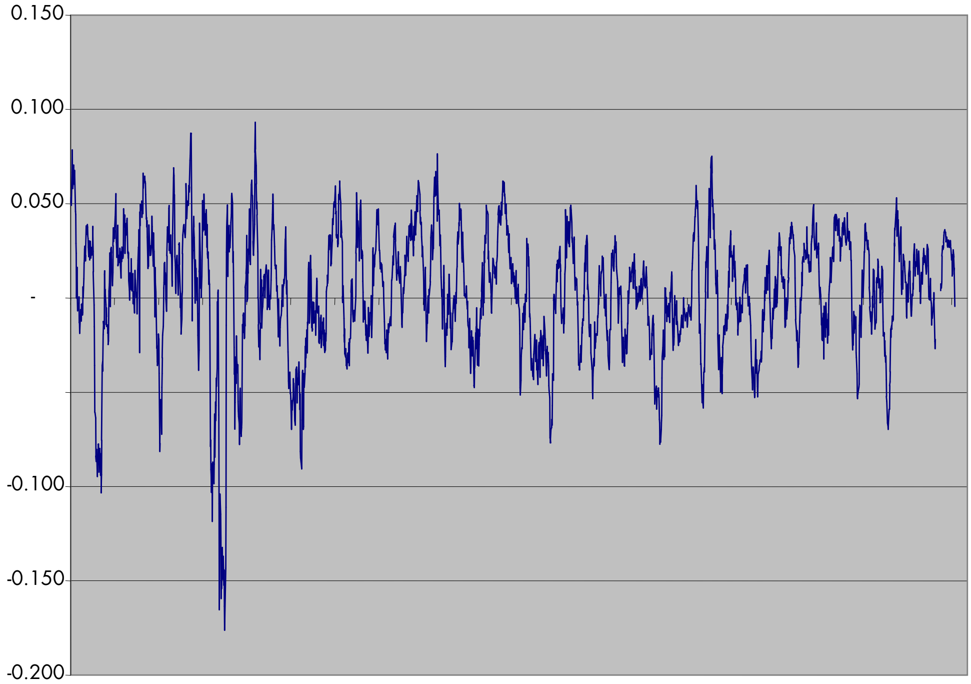


Figure 2.5: Monthly returns of a carry trade. Percentage monthly data from Jan 1 1997 to Jan 1 2007 where bid ask spreads were considered. *Source: M. Salvi. ZKB*

option and the stock one derives the bond property. A portfolio with this property is **hedge position**. The **same** value for A follows as under the replication approach. One could equally take the last combination - the derivative and the bond - as a portfolio combination an replicate the stock.

Assume that the market consists of two risky assets and the option. What happens if the two asset payoffs are linearly dependent, i.e. the assets are redundant? Then, the replication system has no solution. Similar if there is only one asset the option cannot be replicated.

We change our market in the initial example as follows: The time T-values of the risky asset are 80 and 105 and the derivatives pays 10 in the upper state and 0 in the lower one. Forming the replication portfolio and solving the equations we get $A = 0.4, B = -29.1$ and $V_0 = 10.9$. This price makes no sense. Why should anybody pay 10.9 for a contract which pays 10 or 0 at time T? The replication portfolio was setup correctly. Therefore something must be wrong in the market structure. We show that the market is not free of arbitrage. To see this, we write the risky asset price moves using the up ('u') and down ('d') notation, i.e. $120 = 100 * u$ and $80 = 100 * d$. Hence, $d = 1.2, u = 0.8$. We further write $r = 1.1$ for the risk less interest factor.

Proposition 2.3.2. *Arbitrage is not possible in the above one period market if and only if*

$$u > r > d \tag{2.22}$$

holds

To explain the result, we note that $u > d$. Suppose $r > u > d$. Then the risk less investment always dominates the risky one - in all possible stated tomorrow. Therefore, go short the risky asset and long the risk less one. In the case $u > d > r$ a similar argument applies. We obtain for the example at the beginning of the section:

$$1.2 > 1.1 > 0.8 \quad \Rightarrow \quad u > r > d \ .$$

The above proposition gives us a simple criterium to check whether no arbitrage is possible or not in a given binomial market.

We consider **risk neutral pricing**. We have seen that historical probabilities or beliefs about the risky asset price dynamics do not matter for fair option pricing in the replication approach. But there is a pricing approach where probabilities matter. These probabilities are different from the empirical or subjective ones. We define for our one period market the quantity q:

$$q = \frac{r-d}{u-d} \ . \tag{2.23}$$

Proposition 2.3.3. *Suppose that there are no arbitrage possibilities. Then q is a probability, the so-called* ***risk neutral probability****.*

To prove this we show $0 \leq q \leq 1$. Since $r - d > 0$ and $u - d > 0$, we have $q > 0$. Assume $q > 1$. This is equivalent to $r - d > u - d$, i.e. $r > u$. This contradicts the assumption of no arbitrage.

If we calculate q in the original setup we get $q = 0.75$. In the variant with an arbitrage opportunity we have $q = 1.25$. The form of the risk neutral probabilities is

$$q = \frac{\text{Return relative to risk free}}{\text{Volatility}} \ .$$

Again, the original probabilities $p, 1-p$ do not matter. We can use the quantity as follows to characterize no arbitrage. The definition of q is equivalent to $qu - qd + d = r$. Multiplying the last equality with S_0 we get (using the notation $S_T^u = S_0 u$)

$$\underbrace{qS_T^u + (1-q)S_T^d}_{\text{Expected value under } Q} = E^Q[S_T] = rS_0 \ .$$

Dividing by r:

$$E^Q[\frac{S_T}{r}] = S_0 \ .$$

In an arbitrage free market, the expected value of discounted risky assets under the risk neutral probability equals today's discounted asset value (note that $S_0/1$ is the discounted value in 0). This generalizes to many period models - the **martingale property for discounted risky assets**. The First Fundamental Theorem of Finance states that converse also holds. The existence of a martingale measure implies no arbitrage. Since martingales have no drift, the expected value of the discounted price process is constant. Given that interest rates in 2012 are close to zero in Switzerland the martingale property for discounted Swiss stocks implies that one expected zero growth for Swiss stocks. This reasoning is not true since the martingale property holds under the risk neutral probability - the real life probability is obtained by adjusting the risk neutral one by the drift of the stock prices. We consider this below in more details.

We **relate** the replicating approach to the risk neutral one. The solution of the general replication equations

$$\begin{aligned} A*S_0*u+B*r &= C^u \\ A*S_0*d+B*r &= C^d \end{aligned} \tag{2.24}$$

is

$$A = \frac{C^u - C^d}{S_0 u - S_0 d} \ , \ B = \frac{C^u\, d - C^d\, u}{dr - ur}$$

A measures the price sensitivity of the derivative given a price change of the underlying risky asset. This sensitivity is called the Delta, i.e. $A = \Delta$. We have at time 0:

$$V_0 = AS_0 + B = \Delta S_0 + B \ . \tag{2.25}$$

Transforming this expression we get after some algebra:

$$\begin{aligned} V_0 &= C_0 = DeltaS_0 + B \qquad (2.26) \\ &= \frac{1}{r}\Big(C^u q + C^d(1-q)\Big) = E^Q[\frac{1}{r}C_T] \ , \end{aligned}$$

i.e. the fair option price is equal to the discounted payoff under the risk neutral probability Q.

Fair derivative prices are expected values of discounted terminal payoffs under the risk neutral probability. The discounted derivative process is a martingale under the risk neutral probability. The existence of a synthetic probability Q leads to an arbitrage free market structure. The objective or empirical probabilities P does not enter derivative pricing formula.

Instead of a market with a risky and a risk less asset, we can consider a more general setup with two risky assets:

$$\begin{aligned} A*S_T^u+B*X_T^u &= C_T^u \\ A*S_T^u+B*X_T^d &= C_T^d \ . \end{aligned} \tag{2.27}$$

with S and X two assets and C the derivative. Using X as numeraire, the discounted option and risky asset S are both martingales. Indeed, it follows that C or the replicating portfolio is a martingale if and only if S is a martingale:

$$\frac{C_0}{X_0} = E^{Q^X}\left[\frac{C_T}{X_T}\right] \iff \frac{S_0}{X_0} = E^{Q^X}\left[\frac{S_T}{X_T}\right] . \tag{2.28}$$

The probability Q^X depends on the choice of the numeraire:

$$q^X = \frac{\frac{S_0}{X_0} - \frac{S_T^d}{X_T^d}}{\frac{S_T^u}{X_T^u} - \frac{S_T^d}{X_T^d}} .$$

We summarize. First, the advantage of relative pricing w.r.t. X is that the probability Q^X is **independent** of the claim or replicating portfolio and the pricing equation (2.28) holds for **all** derivatives C. Second, both the derivative and the relative asset price are martingale measures, i.e. the measure related to the numeraire. Third, we could choose S as a numeraire instead of X. This leads to a new measure Q^S such that X/S and C/S are martingales under this new measure: The choice of a numeraire does not alter the price of the derivative [4]. One can b chose to most convenient for calculations. Fourth, if we would consider absolute pricing (pricing using a general equilibrium model) instead of relative one, the martingale measure for a derivative depends on the specific derivative payoff V_T. This is a main reason why one uses relative no arbitrage pricing in practice much more often than a fully fledged general equilibrium model.

2.3.1 General One Period Model

Using linear algebra the above toy one period model can be generalized to multiple assets and several periods. The linear replication equations are generalized as follows.

Definition 2.3.4. *Consider a one-period model with $K > 1$ states at time T and $N - 1$ risky assets S and a risk less asset B. The price of asset j at time T in state k is given by $S^j(k)$. The payoff matrix $\mathbb{P}_T$ is defined by*

$$\mathbb{P}_T = \begin{pmatrix} B^1(1) & S^2(1) & \cdots & S^N(1) \\ \vdots & \vdots & \ddots & \vdots \\ B^1(k) & S^2(k) & \cdots & S^N(k) \end{pmatrix} . \tag{2.29}$$

A portfolio or a strategy is a vector $\phi = (\phi_1, \ldots, \phi_N)'$ with X' the transpose of X.

In our basic one period model we have two states ($K = 2$), two assets ($N = 2$) and the strategy $\phi = (B, A)$. The matrix $\mathbb{P}_T$ has the dimension $K \times N$. If $\phi_k < 0$, there is

[4] The numeraire has to be a strictly positive random variable or stochastic process.

a short position in the asset k. The payoff or portfolio value at time T in state k is the product of the payoff matrix and the strategy vector, i.e.

$$V_T(k) = \sum_{n=1}^{N} \phi_n S^n(k) \tag{2.30}$$

and the payoff vector is $V_T = \mathbb{P}_T \phi$.

Definition 2.3.5. *A payoff V_T is attainable given a market structure $\mathbb{P}_T$ if a portfolio ϕ exists such that $V_T = \mathbb{P}_T \phi$. The portfolio ϕ is called a replication portfolio.*

Examples:

We rewrite the replication formula (2.27) in matrix form

$$\mathbb{P}_T \phi = C_T \ , \ \mathbb{P}_T = \begin{pmatrix} 1.1 & 120 \\ 1.1 & 80 \end{pmatrix} \ , \ \phi = (A, B)' \ , \ C_T = (20, 0)' \ . \tag{2.31}$$

Consider the following variation of the one period model: There are not 2 but 3 states for the assets at time T. Everything else remains the same. The matrix $\mathbb{P}_T$ reads:

$$\mathbb{P}_T = \begin{pmatrix} 1.1 & 120 \\ 1.1 & 80 \\ 1.1 & x \end{pmatrix}$$

with x a value for the risky asset in the third state. The linear replication equation has then - if the assets are linearly independent - no solution. Replication is therefore not possible. In the other case, we assume that there are still two states but we have 2 risky assets. Then $\mathbb{P}_T$ is of the form

$$\mathbb{P}_T = \begin{pmatrix} 1.1 & 120 & x \\ 1.1 & 80 & y \end{pmatrix} \ .$$

The linear replication equations typically will have an infinite number of solutions. This shows that the number of states and the number of risky determine the market structure. The system

$$V_T = \mathbb{P}_T \phi \ \to \phi = \mathbb{P}_T^{-1} V_T$$

has a solution if $\mathbb{P}_T$ is invertible. If $N = K$ and if the rows or columns are linearly independent the inverse exists. Intuitively, for $N = K$ the sources of randomness, i.e. the number of risky assets, can spanned in all state by the assets. What happens if there are more states than assets, i.e. $K > N$? Then the replication problem has no solution. In the case $K < N$ there are more variables ϕ than equations. Linear algebra implies that an infinite number of solutions is the generic case. We have fewer risk sources than assets.

Definition 2.3.6. *A market with a payoff matrix $\mathbb{P}_T$ is complete if each claim V_T can be obtained, i.e. there exists a strategy such that $V_T = \mathbb{P}_T \phi$.*

If $\mathbb{P}_T$ is invertible, market completeness follows. This means that the rank of $\mathbb{P}_T$ has to be equal to K, i.e. the number of assets is not lower than the number of states. Given a market with payoff matrix $\mathbb{P}_T$ and asset price vector S_0, arbitrage is defined as follow:

Definition 2.3.7. *Consider a market with payoff matrix $\mathbb{P}_T$ and asset price vector S_0. An arbitrage is a portfolio $\phi = (\phi_1, \ldots, \phi_N)'$ such that*

- *the initial portfolio value $V_0 = \langle S_0, \phi \rangle \leq 0$[5] is not positive,*
- *the payoff of the portfolio ϕ is not negative for all states k and*
- *and there exists at least a single state $\tilde{k}$ where $V_T(\tilde{k}) > 0$ or $\langle S_0, \phi \rangle < 0$ holds.*

Starting with no money the strategy leads to a portfolio where no loss is possible and if some specific states are realized even a gain follows. When do we know that a market is free of arbitrage? In the one period model the statement of no arbitrage can be restated equivalently in terms of geometry and linear algebra. An application of the Separating Hyperplane Theorem leads to the following characterization of no arbitrage:

Proposition 2.3.8 (First Fundamental Theorem of Finance (FFTF))**.** *There is no arbitrage opportunity if and only if there exists a vector $\psi \in \mathbb{R}^K, \psi_j > 0$ for all j, such that*

$$\mathbb{P}_T' \psi = S_0 \ . \tag{2.32}$$

The proof is given in Appendix 7.3.

This theorem is a characterization theorem. It characterizes a relationship that must exist between given **security prices** and **payoffs** if no arbitrage opportunities are to exist. It does not tell what security prices should be. They come from market equilibrium (i.e., supply = demand). The FFTF states that the price of asset i at time 0 is given by

$$S^i(0) = \sum_{j=1}^{K} \psi_j S^i(j) \ .$$

To further specify and obtain an interpretation, consider first the risk less asset, i.e. $i = 1$, Then

$$B(0) =: B_0 = \sum_{j=1}^{K} \psi_j B^1(j) = \sum_{j=1}^{K} \psi_j \times 1 = \sum_{j=1}^{K} \psi_j$$

since the risk less asset pays 1 in all possible K states. If we normalize the right hand side of the last equation, i.e. we set

$$\phi_0 := \sum_{j=1}^{K} \psi_j$$

[5] We write $\langle x, y \rangle$ for the scalar product of two vectors x, y.

and

$$q_i = \psi_i / \psi_0 \ , \quad \forall i \ ,$$

the risk less asset can be rewritten

$$B_0/\psi_0 = \sum_{j=1}^{K} q_j B^1(j) = 1 \ .$$

Therefore, ψ_0 is the discount on a risk less borrowing. If r in the risk less annual interest rate, we write

$$B_0 = \psi_0 = \frac{1}{(1+r)^T} \ .$$

This implies for all other risky assets:

$$S^i(0) = \sum_{j=1}^{K} \psi_j S^i(j) = \frac{1}{(1+r)^T} \sum_{j=1}^{K} q_j S^i(j) = E^Q[\frac{S^i}{(1+r)^T}] \ .$$

Since the sum of the q_i's is equal to 1 and for all $q_i > 0$ - the q's are probabilities. The q_i are called risk neutral probabilities. This shows the equivalence of risk neutral probabilities and state prices:

$$\frac{q_i}{(1+r)^T} \leftrightsquigarrow \psi_i \ .$$

To obtain an interpretation of the state price vector, we introduce the Arrow-Debreu securities, that is a set of securities $e(j), j = 1, \ldots, K$ where the security $e(m)$ pays 1 CHF if the state m is realized and zero else. The First Fundamental Theorem of Finance implies

$$\begin{pmatrix} 1 \\ e(1) \\ \cdots \\ e(K) \end{pmatrix} = \begin{pmatrix} (1+r)^T & \cdots & \cdots & (1+r)^T \\ 1 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_1 \\ \cdots \\ \psi_K \end{pmatrix} \ .$$

Hence, $e(j) = \psi_j$ follows.

Corollary 2.3.9. *State-price densities are the prices of the Arrow-Debreu securities.*

Using the FFTF the stochastic discount factor D satisfies

$$S_0^j = E[D(0,T)\mathbb{P}_j]$$

with $\mathbb{P}_j$ the j-th row of the payoff matrix and the expectation w.r.t. to the objective or statistical probability measure P. The stochastic discount factor in state s, D_s is equal to $D_s = \psi_s / p_s$, i.e. it is the state price, per unit probability.

We state and prove the Riesz-Fischer Theorem in Appendix 7.3. This theorem is the basis for the above **linear pricing** relationship. It states that any linear function on a particular vector space has an scalar product representation.To apply this to our case: A system of prices (S_0, D) is consistent with no arbitrage if and only if there exists a SDF D such that the linear representation $S_0^j = E[D(0,T)\mathbb{P}_j]$ holds for all j. The expected value has the property of a scalar product, i.e.

$$S_0^j = E[D(0,T)\mathbb{P}_j] = \langle D(0,T), \mathbb{P}_j \rangle \ .$$

Hence the Riesz-Fischer Theorem states that if $f(\mathbb{P}_j)$ is linear then it f can be uniquely represented by the scalar product, i.e. the expectation.

Given the equivalence between state-price densities and risk neutral probabilities, we restate the First Fundamental Theorem of Finance:

Proposition 2.3.10 (First Fundamental Theorem of Finance)**.** *There is no arbitrage opportunity if and only if a risk neutral probability exists.*

The absence of no arbitrage does not implies that the risk neutral probability is unique. In fact, there can be an infinite number of them. Which one should we take for pricing? Our market structure together with the no arbitrage condition is not sufficient to imply an unique price. Such markets are called **incomplete**. The Second Fundamental Theorem of Finance considers the case where pricing is unique.

Proposition 2.3.11 (Second Fundamental Theorem of Finance)**.** *Consider an arbitrage free market. The state price vector is unique if and only if the market is complete.*

Hence, market completeness is equivalent to the uniqueness of the risk neutral probability.

Proof. To show that completeness implies uniqueness of the state vector, assume that there exist two state vectors ψ_1, ψ_2 which both solve the equation $S_0 = \mathbb{P}\psi$. This then implies that $0 = \mathbb{P}(\psi_1 - \psi_2)$, i.e. the vector difference is orthogonal to all rows of the payoff matrix. Therefore, the difference is not attainable which contradicts that the two vectors are state price vectors. The other direction is proved with a similar argument. □

Examples

1. Consider the following structure:

$$\mathbb{P}_T = \begin{pmatrix} 4 & 6 & 2 \\ 12 & 3 & 9 \end{pmatrix} \ , \ S_0' = (7, 3, 5)$$

The equation for the state price density reads

$$D_T' \psi = S(0) \ .$$

One solution of the linear system

$$\psi' = (1/4, 1/1) \ .$$

The market is free of arbitrage. That the system has a solution is the exception and not the rule. That is, a system of 3 equations and 2 unknowns typically has no solution. Intuitively we have 3 securities and 2 states. The expectation is that the payoffs of the three non-redundant securities are conflicting, i.e. the system has no solution and arbitrage is possible.

2. We consider the following market with a risk less and a risky asset with 3 states. This is the next two examples are taken from Frey and Schmidt (2006)

$$\mathbb{P}_T = \begin{pmatrix} 1 & 180 \\ 1 & 150 \\ 1 & 120 \end{pmatrix} \ , \ S(0) = (1, 150)'$$

i.e. we assume that interest rates are zero. The defining equation $D'\psi = S_0$ reads explicitly

$$\begin{aligned} 150 &= 180\psi_1 + 150\psi_2 + 120\psi_3 \\ 1 &= \psi_1 + \psi_2 + \psi_3 \ . \end{aligned}$$

Since there are more unknowns than equations, one expects a continuum of state price densities to satisfy the no arbitrage condition. The solution reads

$$\psi_1 = \psi_3 \ , \ \psi_2 = 1 - \psi_3 \ .$$

Since all state prices have to be strictly positive, the following parametrization describes the set of all state prices:

$$\psi = \{(a, 1-2a, a) \ , \ a \in (0, 1/2)\} \ .$$

This market is free of arbitrage within the given parametrization set but incomplete. This means that there exist self-financing portfolios ϕ such that there are claims V_T which are not replicable in all possible future states at date T: Hence, $V_T \neq \mathbb{P}_T\phi$ for some states. In this example consider a call option which payoff $(30, 0, 0)$. The standard replication equations are

$$\begin{aligned} A + B * 180 &= 30 \\ A + B * 150 &= 0 \\ A + B * 120 &= 0 \ . \end{aligned}$$

The first and third equation imply $A = -60, B = 1/2$. But this choice does not solve the second equation. Hence the call option is not attainable.

Since the difference $V_T - \mathbb{P}_T\phi$ is not zero, **hedge risks** exist. No arbitrage alone does not leads to a unique price a second criterion is needed to enforce uniqueness. The most important one is the market, i.e. the price of a derivative depends on

parameters which are not fully specified within the model itself. The derivative price is fixed by mapping the parameterized prices to observed market prices. This approach is used in interest rate modelling for example and is known as "inverting the yield curve". Academic research considers various other approaches such super-replication, quadratic hedging, quantile hedging for example.
Super-replication is e a criterion where one searches for a strategy ϕ such that $\mathbb{P}_T\phi \geq V_T$ holds and the strategy costs are minimal, i.e. $\langle S_0, \phi\rangle$ attains a minimum value. Using such a strategy, the seller of the product is on the safe side ($\mathbb{P}_T\phi \geq V_T$). The price of this strategy is often prohibitive large. This is due to the requirement that the hedge portfolio is in all possible states worth at least the liability V_T. It follows that super replication amounts to solve the problem

$$\sup\{E^Q[\tilde{V}_T] \ , \ Q \text{ a risk neutral probability}\} \ ,$$

where $\tilde{V}$ is the discounted payoff. Hence, super replication is to find the risk neutral probability which gives the highest discounted payoff value. If we introduce the claim $V_T = (30, 0, 0)'$ in our example, we have to solve

$$\sup\{a30 + (1 - 2a)0 + a0, \ a \in (0, 1/2)\} \ .$$

The solution is given by $a = 1/2$ which implies

$$\sup\{E^Q[\tilde{V}_T] \ , \ Q \text{ a risk neutral probability}\} = 15 \ .$$

This leads to a degenerate state price vector $\psi = (1/2, 0, 1/2)$, i.e. a corner solution where at least one state price density is zero. Since state price densities have to be strictly positive such that arbitrage is not possible the calculations in incomplete markets deliver prices which form the bounds for no arbitrage prices. Using this vector one replicates the claim

$$\phi_1' + 180\phi_2' = 30 \ , \ \phi_1' + 120\phi_2' = 0 \ .$$

The solution is $\phi_1' = -60$ and $\phi_2' = 1/2$. This leads to the costs

$$\langle S_0, \phi'\rangle = -60 + 1/2 \times 150 = 15 \ .$$

There the fair premium equals the upper bound of the expected value over all risk neutral probabilities.

3. We consider an incomplete market in one period with two securities S (risky) and B (risk less). The security B pays in each state at the end of the period CHF 1. The risky security S can achieve three states at the end of the period: $S^u = S_0u > S^m = S_0m > S_d = S_0d$ with S_0 the initial price and the up/mid/down parameters u/m/d. This model extends the basic complete market binomial or Cox-Ross-Rubinstein model, see Section ??. The trinomial model is incomplete since the payoff matrix has the dimension 3×2 or alternatively, there are more

states than traded asset to span the states. Since there is a risk less asset, the three state prices add up to the risk less discounting factor:

$$\frac{1}{1+R} = \psi_1 + \psi_2 + \psi_3 \ . \tag{2.33}$$

The second condition $S(0) = \sum_j \psi_j S^j$ follows from no arbitrage. Since $S^j = S(0) \times x$ with $x = u, m$ or d the condition reads:

$$1 = u\psi_1 + m\psi_2 + d\psi_3 \ . \tag{2.34}$$

Equation (2.33) and (2.34) are two equations for the three dimensional state price vector. In the binomial models there are two equations for a two dimensional state price vector. A unique state price vector exists. The market is complete. In the trinomial model the solution of the two equations is in general a line which shows that there is a continuum of no arbitrage free prices, see Figure 2.6.

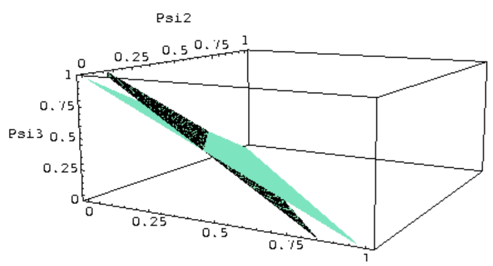


Figure 2.6: Intersection of the two planes (2.33) and (2.34) defining the bounded line of arbitrage free prices. The parameters are $u = 1.2, m = 1.1, d = 0.9, R = 0.4$.

Therefore, the state price vector is not unique. Despite the incompleteness the no arbitrage condition is the same as in the binomial model: There is no arbitrage if and only if

$$d < 1 + R < u \ .$$

The line is bounded by the requirement that state prices are positive. Each vector on the line segment used to price derivatives leads to arbitrage free prices. Solving two equations, the boundary points of the line segment follow: One gets for $m \geq 1 + R$

$$\psi_1 = \frac{1+R-d}{(1+R)(u-d)} \ , \ \psi_2 = 0 \ , \ \psi_3 = \frac{u-1-R}{(1+R)(u-d)} \ ,$$

and

$$\psi_1 = 0 \ , \ \psi_2 = \frac{m-1-R}{(1+r)(m-d)} \ , \ \psi_3 = \frac{1+R-d}{(1+R)(m-d)} \ .$$

A similar solution holds it the mid-move m is lower than the risk less one ($m < 1+R$). The boundary values do not lead to arbitrage free prices since some components of the state price densities are not strictly positive. It is a general fact, that extreme or boundary values lead to arbitrage opportunities. If $m \to d$ or $m \to u$, the trinomial model collapses to the binomial one, the state price are

$$\psi_1 = \frac{1}{1+R}q \ , \ \psi_2 = \frac{1}{1+R}(1-q) \ , \ \psi_3 = 0 \ .$$

Consider a call option in this market with strike $S^m < K < S^u$. The cash flows are $S^u - K$ in state 1 and zero in the two other states. The no arbitrage price is the expected value

$$C = \psi_1(S^u - K)$$

Using the two possible state prices

$$\psi_1 = \frac{1+R-d}{(1+R)(u-d)} \ , \ \psi_1 = 0$$

in the case $m \geq 1+R$ the upper C^+ and lower bound C^- for the call price follow:

$$C^+ = \frac{1+R-d}{(1+R)(u-d)}(S^u - K) \ , \ C^- = 0$$

For $m < 1+R$, $\psi_1 = \frac{1+R-m}{(1+R)(u-m)}$, holds and the option price is

$$C = \frac{1+R-m}{(1+R)(u-m)}(S^u - K) \ .$$

If the trinomial model tends to a binomial one, i.e. $m\ tod$ for example, the upper and lower bond prices collapse, i.e.

$$C^+ = C = \frac{1}{1+r}q(S^u - K) \ .$$

Before we do the next step in option pricing, the generalization to multi periods, we make a break and consider other topics of options in the next section.

2.3.2 Options Basics and Option Strategies

We considered the pricing of options in the last sections. But there is more than pricing. First, there are properties of options which are pricing model independent. Second, options can be used to define investment strategies. We introduce some basic option types in this section, discuss model-free relationships and propose some classic, static option strategies.

A European-style call option/warrant confers the right to purchase at a specific point (maturity T) in time a specific amount (strike K) of a specific underlying security (S) at a specific price. If the right to purchase the underlying is given for a whole period the option is of the American style. Call warrants are essentially the securities form of standardized call options. Consequently, they are easier to trade for private investors. If in the definition the right to buy is changed to the right of sell the option is called a put option.

Options can be classified according different characteristics: The exercise-type (European or American), the complexity of the payoff (simple or vanilla and exotic) and the trading type (warrant, OTC or standardized options such as Eurex-options).

The option price has two components: the intrinsic value, i.e. the value when you sell an option at a given date, and the time value. The intrinsic value of a call with strike K is

$$C_t = \max(S_t - K, 0) =: (S_t - K)^+$$

with S_t the price of the underlying at time t. This is the value when the call is exercised at a date t. If $S_t - K > 0$, the option is in-the-money and it pays to exercise. If $S_t - K < 0$ at any date t it is not rational to buy the underlying at a price K is it worth the lower value S_t - hence one uses the optionality of not exercising the call. This leads to the kink or maximum in the payoff formula. At maturity T the intrinsic value equals the payoff and time value is zero. The intrinsic value can never be negative since its the intrinsic value of an option and not of an obligation. This value is model independent, i.e. it can be read off from market data. What can be said about the time value? We consider at a date $t < T$ the portfolio V_t in Table 2.1:

Portfolio V	V_t	V_T if $S_T < K$	V_T if $S_T \geq K$
Short S	$-S_t$	$-S_T$	$-S_T$
Long call C	$+C_t$	0	$S_T - K$
Risk less	$PV(K)$	$+K$	$+K$
Portfolio value	$-S_t + C_t + PV(K)$	$K - S_T > 0$	0

Table 2.1: Time value of a call.

We assume in this first part that the underlying stock is paying no dividends. The portfolio value in 2.1 is at maturity never negative. No arbitrage implies that the portfolio can never be negative at any prior date t too, i.e.

$$-S_t + C_t + PV(K) \geq 0 \ .$$

Assuming positive interest rates, $PV(K)$ is smaller than K. Therefore,

$$C_t \geq S_t - PV(K) > S_t - K \ ,$$

that is the option price at any time t prior to maturity is strictly larger then the intrinsic value. Therefore, the call option price for positive interest rates is never smaller than the intrinsic value - the difference is the time value. The non linearity of the payoff for options is basic to generate a time value. To see this, we linearize the payoff in 2.2, i.e. at the kink of the strike value we would have the following cash flows (bold faced) Therefore

Portfolio V	V_t	V_T if $S_T < K$	V_T if $S_T \geq K$
Short S	$-S_t$	$-S_T$	$-S_T$
Long derivative C'	$+C'_t$	$S_T - K$	$S_T - K$
Risk less	$PV(K)$	$+K$	$+K$
Portfolio value	$-S_t + C'_t + PV(K)$	0	0

Table 2.2: Time value of a linear payoff.

the time value of such a linear payoff, i.e. a forward contract, is zero. Contrary to the intrinsic value the quantification of the time value requires a model.

Options satisfy structural constraints which a pricing model independent, see Figure 2.7.

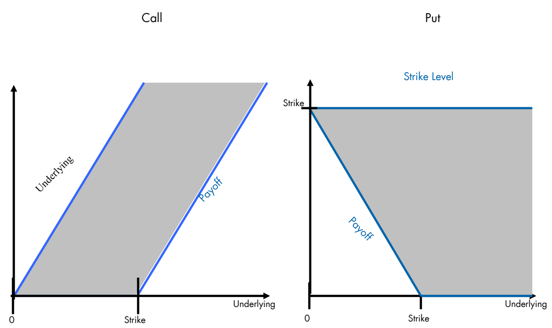


Figure 2.7: Model independent price bounds for call options (European or American style) and American puts. *Source: Zurich Cantonal Bank.*

The value of a call is always less or equal than the underlying value. To understand this, assume a strike of 0, then the call value is equal to the underlying value. For any higher strike the call value is lower than the underlying value. The values are bounded above by the intrinsic value - the price bands for a call follows. For puts similar a reasoning implies that the put value is smaller or equal to the strike. If C and C_a are the call prices for an European and American call, respectively, the first restriction states that all prices are non-negative which reflects the optionality. A second restriction is that at maturity the American and European style options have the same value since time value is zero and the American feature to early exercise is worthless. Third, since American options can be exercised any time they must always be worth at least their intrinsic value, else arbitrage is possible by buying the option and immediate exercising them. This condition is not valid for European style options. Fourth, since additional rights cannot have negative value the American style options cannot be worth less than their European style counter parts. For non-dividend paying stocks the price of an American call is **equal** to the price of an European one. This is surprising since the American call seems to have a more valuable optionality than the European one. To understand the equality, we recall that the value of an European option is not less than its intrinsic value. Assume that it is worth to early exercise at $t < T$. One obtains the underlying value S_t and pays the strike K. Hence one gives up the interest rate on K between t and T. Therefore, it is optimal to pay K as late as possible, i.e. at maturity T. This shows that early exercise is not optimal which sets the price of the American call equal to the European one. If the stock pays dividends, to sell a call early makes sense if the received dividend value exceeds the described interest rate loss. For puts, even without dividends, we always have that $P_a \geq P$. Consider a put with strike $K = 10$ and assume that the underlying becomes worthless. Early exercising gives a profit of 10 and waiting does not increase the profit. Even worse, 10 today are worth more than 10 in the future. Therefore early exercise can be rationale for puts. The structural difference between calls and puts is that for former one pays for the strike whereas for puts one obtains the strike.

Fifth, the higher the strike for a call the lower the call price since profitability or the change to be ITM (in-the-money, i.e. $S > K$) decreases with increasing strike. The opposite is true for puts. Sixth, a similar property holds for different time to maturities. Since an additional time period can do no harm the price of a call or put should be not decrease with an increasing time to maturity. Seventh, since the underlying can be seen as a call option with strike 0 and infinite maturity it follows together with the facts under four and five that the price of the stock must be at least as high as the price of an American call. Therefore, the stock price is an upper price boundary for call options - American or European style ones. Eights, since the stock is an asset with limited liability the American put values have the strike value as upper boundary value.

We consider **break-even** and **leverage**. Consider a stock with price $S_0 = 27.6$ in a currency, a put option with strike $K = 24.5$, a option price $P_0 = 0.125$ and a ratio of 1:2, i.e. two options are necessary to buy 1 stock. **Break-even**, i.e. the profit is zero at maturity, is defined by the equation

$$P_T = 0.125 = \max(K - S_T) \times \text{Ratio} .$$

Therefore to make no loss the stock S_T is allowed to attain at most the value

$$S_T = K - \frac{P_T}{\text{Ratio}} = 24.25 .$$

Leverage considers the Delta of the option, i.e. it is defined as:

$$\text{Leverage} = \frac{\text{Price Underlying}}{\text{Price Option x Ratio}} \times \Delta = \frac{S}{C \times R} \frac{\partial C}{\partial S} .$$

This For an investment in a stock, Leverage is 1. Consider a call with price 10, Delta of 0.75, Ratio of 1:1 and price of the stock of 22.50. Leverage is then 3: The call option offers a three time higher participation at the stock compared to a direct stock investment.

Put-call parity is an important **model-independent** relationship for European style options. The parity states that a put is a call and vice versa. We consider the case of non-dividend paying stocks. The parity is essentially the mathematical equality

$$\max(S - K, 0) - \max(K - S, 0) = S - K .$$

In more finance terms, form a portfolio V (i) short a call, (ii) long a put, (iii) long a stock and (iv) short cash. The European call and put have the same maturity, strike and underlying. Then V has in both states $S_T > K$ and $S_T \leq K$ **value zero at time** T. No arbitrage implies that $V_0 = 0$ has to hold, i.e. the put-call parity follows:

$$C_t - P_t = S_t - PV_t(K) . \tag{2.35}$$

If the stock pays known dividends D, the value of the stock is reduced by the PV of the dividends. The put-call parity then reads:

$$C_t - P_t = S_t - PV_t(K) - PV_t(D) . \tag{2.36}$$

To derive the parity for futures we note that $F(t,T) = S(t)e^{-r(T-r)}$ or we set up two portfolios:[6]

$$C_t - P_t = (F(t,T) - K)e^{-r\tau} . \tag{2.37}$$

A similar argument gives the put-call parity for FX underlyings. We get

$$C_t - P_t = S_t e^{-r_F T} - K e^{r_D t} \tag{2.38}$$

with r_D the domestic and r_F the foreign risk free rate.

Examples:

- Consider a call on 6m silver futures with price CHF 0.56 per ounce and strike 8.5 CHF. The silver spot price is CHF 8.0 and risk less rate is 10%. The price of the put option with the same characteristic is given by

$$0.56 + 8.5e^{-0.5 \times 0.1} - 8e^{-0.5 \times 0.1} = 1.04 \text{ CHF} .$$

[6] Portfolio V is long a call on the future F with strike K and long cash $Ke^{-r\tau}$. We have

$$V_T = \max(F(T) - K, 0) + Ke^{-r0} = \max(F(T), K) .$$

Portfolio W is long a put on the same future, long the future and long cash $Fe^{-r\tau}$. It follows, that

$$W_T = V_T$$

and by no arbitrage the two portfolios have to be equal for all times which implies the parity relation.

- Consider next a European call on USDCHF underlying S, i.e. the underlying expresses how much CHF equals one USD. The risk free rate in the U.S. is 2 percent and 1 percent in Switzerland. The spot is at 1.05, maturity of the call is 1 year and the 1.10call price, i.e. the call with strike $K = 1.10$, is 0.025 CHF per 1 USD. What is the 1.10put price? We get $P = 0.0141$ CHF.

- Put-call parity implies for positive interest rates:

$$\begin{aligned} C &= S - PV(K) + P = S - K + K - PV(K) + P \\ &= \underbrace{S - K}_{\text{Intrinsic Value}} + \underbrace{\underbrace{(1 - e^{-rt})K}_{>0} + \underbrace{P}_{>0}}_{>0} . \end{aligned}$$

Since the price of an American is worth not less than an European one, early exercising means to throw away a put option. If the stock pays dividends, the above analysis reads

$$\begin{aligned} C &= S - PV(K) + P - PV(D) = S - K + K - PV(K) + P - PV(D) \\ &= \underbrace{S - K}_{\text{Intrinsic Value}} + \underbrace{\underbrace{\underbrace{(1 - e^{rt})K}_{>0} + \underbrace{P}_{>0}}_{>0} - PV(D)}_{?} , \end{aligned}$$

i.e. early exercise is rational if

$$(1 - e^{rt})K + P - PV(D) < 0$$

holds: The interest lost by early exercising has to be smaller than the gained dividend.

2.3.3 Basic Option Strategies

An option strategy is a portfolio of options with a single or several objectives. Possible objectives are:

- Hedging. Reduce risks in an existing portfolio of securities.
- Investing. Customized risk and return profile.
- To take advantage of mispriced securities, i.e. arbitrage strategies.

A characteristic of an option strategy is whether the strategy is **static**, i.e. the strategy is setup at initiation and left unchanged over time until maturity, or whether the strategy is **dynamic**. Dynamic strategies can be rule based or discretionary or a mixture of them.

For call and put strategies the loss/gain is

- bounded by the premium for long call and long put,

- bounded by the exercise price minus the premium for short put,
- unbounded for short call.

The gain potential is

- bounded by the premium for short put and short call,
- bounded by the exercise price minus the premium for long put,
- unbounded for long call.

The next more complex strategies are those consisting of an option and the underlying value. We consider:

- Covered call (investment strategy), see Figure 2.8;
- Protective put (hedge strategy).

The investment motivation for a covered call is a belief that the underlying value will move sideways or at most increase only moderately. The investor does not search protection for a stress or a crash scenario. The covered call strategy, i.e. long the underlying and short a call option, capture this investment motivation. If the underlying decreases the value of the covered call is always higher than the underlying values since the investor obtained a premium for the sold call option.

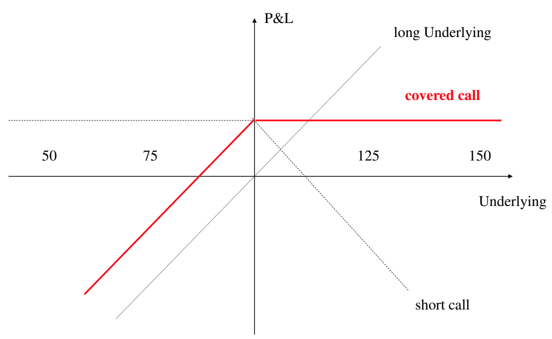


Figure 2.8: Covered Call.

Table ?? shows: The covered call is always worth more than the underlying if it drops, there is no capital protection and the upside is given up if above a certain level. If the motivation of the investor is to buy protection he goes long the underlying and long a put. This strategy - protective put - reduces loss potential to the put premium keeping the gain potential unbounded.

S_T	Payoff Call	P&L Short Call	P&L Long S	P&L Covered Call
27	0	3.6	-4	-0.4
28	0	3.6	-3	0.6
29	0	3.6	-2	1.6
30	0	3.6	-1	2.6
31	0	3.6	0	3.6
32	0	3.6	1	4.6
33	0	3.6	2	5.6
34	-1	2.6	3	5.6
35	-2	1.6	4	5.6
36	-3	0.6	5	5.6

Table 2.3: P&L-comparison for covered call, long underlying, short call for the data $K = 33$, $S(0) = 31$. The upside is restricted to $5.6 = K - S(0) + C(0) = 33 - 31 + 3.6$.

The next strategy is the so-called **3-options strategy**. Consider 3 European or American call options with increasing strikes $K_1 < K_2 < K_3$ with the same underlying and maturity. If arbitrage is not possible, the following structural,i.e. model-independent, inequality holds:

$$C(K_2) \leq \frac{K_3 - K_2}{K_3 - K_1} C(K_1) + \frac{K_2 - K_1}{K_3 - K_1} C(K_3) \ . \tag{2.39}$$

It exists a combination of strikes such that the option price to the middle strike K_2 is never larger than a combination of the minimum price $C(K_1)$ and the maximum price $C(K_3)$. This relation is due to the fact that option prices are monotone decreasing with increasing strikes.

We show how the above 3-option relation can be used to test for arbitrage. Assume $S_0 = 100$ and three American calls with strikes and prices:

- $C_1(94) = 8.4$.
- $C_2(100) = 5.5$.
- $C_3(108) = 1.4$.

Are these prices arbitrage free? To answer this, we write the middle strike $K_2 = 100$ as a convex combination of the two others:

$$K_2 = 100 = a * K_1 + (1 - a)K_3 = a * 94 + (1 - a)108 = 108 - 14a,$$

$a = 4/7 = \frac{K_3 - K_2}{K_3 - K_1}$ and $1 - a = 3/7 = \frac{K_2 - K_1}{K_3 - K_1}$ follow. Using this, we check (2.39):

$$\underbrace{\frac{K_3 - K_2}{K_3 - K_1}}_{a} C(K_1) + \underbrace{\frac{K_2 - K_1}{K_3 - K_1}}_{1-a} C(K_3) = 4/7 * 8.4 + 3/7 * 1.4 = 5.4 \ .$$

This price is smaller than $C(K_2) = 5.5$. Arbitrage is possible. The K_2-call is relative to the two other ones to expensive. Can you exploit this?

The next strategy, called **conversion**, shows how no arbitrage violations measured in the put-call parity can be exploited. We recall the put-call parity for futures:

$$C - P = (F - K)e^{-r\tau}$$

and consider the example strike of call and put $K = 110$, futures price $F = 100$, put price $P = 12$, call price $C = 4$, $r = 10\%$ and time to maturity $\tau = 1$ year. Inserting this into the parity gives.

$$C - P = -8 \ \neq \ (F - K)e^{-r\tau} = -9.05$$

Since the call is out of the money and the put is in the money it is more likely that the call is mispriced. To exploit the arbitrage opportunity we buy the low priced put and sell the high priced call. At time t we setup the portfolio V:

- Long Put, i.e. $P = -12$.
- Short Call, i.e. $C = 4$.
- Long den Future, i.e. $F = 0$.
- Borrow Cash to finance the strategy, i.e. 8.

This portfolio value is zero at time t. What is the value at time T?

$$\begin{aligned} V(T) &= P_T - C_T + F - Xe^{r\tau} \\ &= \underbrace{(K - S_T)^+}_{\text{Put}} - \underbrace{(S_T - K)^+}_{\text{Call}} + \underbrace{(S_T - F)}_{\text{Future}} - \underbrace{Xe^{r\tau}}_{\text{Pay back loan}} \end{aligned}$$

The value of the long futures equals the value difference between the underlying at maturity S_T and the contracted futures price F at time t. This gives two portfolio values at T:

- $S_T \geq K$

$$\begin{aligned} V(T) &= P_T - C_T + (S_T - F) - Xe^{r\tau} \\ &= 0 - (S_T - K) + (S_T - F) - Xe^{r\tau} \\ &= 0 - S_T + 110 + S_T - 100 - 8e^{0.1} \\ &= 110 - 100 - 8.84 = 1.16 \end{aligned}$$

- $S_T < K$

$$\begin{aligned} V(T) &= P_T - C_T + (S_T - F) - Xe^{r\tau} \\ &= (K - S_T) - 0 + (S_T - F) - Xe^{r\tau} \\ &= 110 - S_T - 0 + S_T - 100 - 8e^{0.1} \\ &= 110 - 100 - 8.84 = 1.16 \end{aligned}$$

An arbitrage strategy is found: A strategy which is worth zero at time t and which is worth 1.16 at time T in all possible states. We note that this risk less gain equals the future value of the initial mispricing in the put-call parity:

$$(9.05 - 8)e^{0.1} = 1.16 \ .$$

The next strategies are **spread strategies**. They are generated by at least two options where the options are identical expect in one or possible two parameters. Variations in strikes are bull and bear spreads, vertical spreads or risk reversals; variation in maturity are calendar or horizontal spreads; variations in the option right are straddles and variations in the option right and strike lead to strangles and butterflies.

For a bull spread, the investor believes that the underlying will increase to a certain level and he wants to restrict losses if the belief turns out to be wrong. He invests in a bull spread, i.e. a long call $C(K_L)$ and short call $C(K_H)$:

$$\text{Bull Spread} = C(K_L) - C(K_H), \text{ where } K_H > K_L$$

The loss is restricted to the difference in the premia, see Figure 2.9.

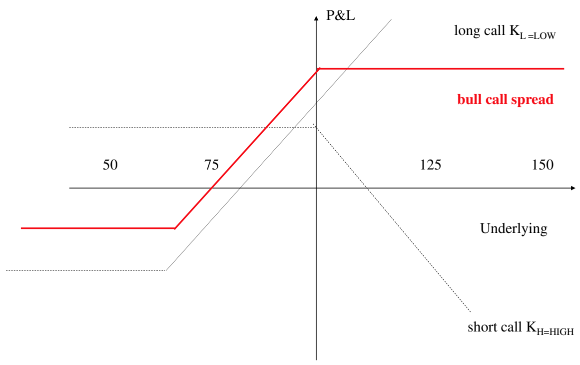


Figure 2.9: Bull Spread.

Verify the following figures for the **Bull Spread:**

$$\begin{aligned} \text{P\&L} &= \max(S(T) - K_L, 0) - \max(S(T) - K_H, 0) - C_L(0) + C_H(0) \\ \text{max Profit} &= K_H - K_L - C_L(0) + C_H(0) \\ \text{max Loss} &= C_H(0) - C_L(0) \\ \text{Break even} &= K_L + C_L(0) - C_H(0) \end{aligned}$$

If an investor beliefs that events will move the underlying away from its present price but he has no directional view, strangle or straddle allow to invest into this volatility bet, see Figure **??**.

A strangle has contrary to the straddle two options with different strikes. A strangle is cheaper than the comparable straddle.

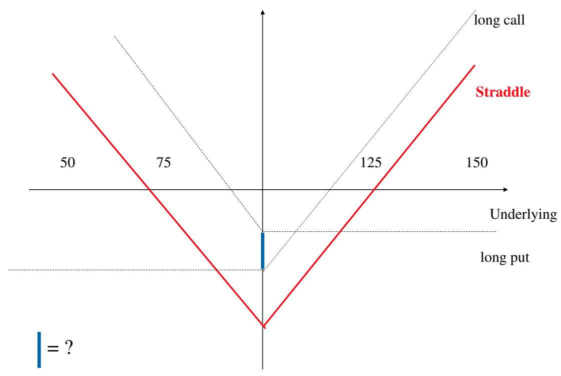


Figure 2.10: Straddle.

2.4 No Arbitrage in a Multi Period Setup

2.4.1 Cox-Ross-Rubinstein (CRR) Model

The ideas of the one period option pricing model transfer to the multi period modelling: The existence of a martingale measure is equivalent to the absence of arbitrage and the uniqueness of such a measure defines a complete market. We discuss the standard Cox-Ross-Rubinstein (CRR) model.

The CRR is **a discrete time, discrete state model** with **two assets**: a risk less B and a risky asset S. The risk less interest rate describing the risk less asset is $r > 0$ and the dynamics of the risk less asset is $B_t = (1+r)^t$ with $B_0 = 1$. The initial price of the risky asset is S_0 and we define the dynamics of the risky asset as follows: The price in period $t+1$ can go up (u) or down (d) with constant rates d and u starting from the t-price. That is, the dynamics reads under an objective probability measure $P = (p, 1-p)$

$$S_{t+1} = \begin{cases} S_t(1+u), & \text{with probability } p; \\ S_t(1+d), & \text{with probability } 1-p. \end{cases} \quad t = 0, 1, \ldots, T-1.$$

The risky asset can be represented by a recombining binomial tree. The price at time k is then

$$S_k = S_0(1+u)^{N_k}(1+d)^{k-N_k}$$

with N_k the number of upwards moves in k time steps.
The values of a portfolio V are given by

$$V_t = \phi_t B_t + \psi_t S_t \ , \ \ \Delta V_t = \phi_t \Delta B_t + \psi_t \Delta S_t$$

where $\Delta V_t = V_t - V_{t-1}$, ϕ_t the amount of CHF invested in risk less asset at time t and ψ_t the number of shares held at time t. The number of shares ψ_t has to be known before

t, that is a time $t-1$. This property of random variable sequences (stochastic processes) is called **predictability**.

We only consider **self-financing** strategies. Assume that the portfolio value V is given by a single asset $V_t = \phi_t X_t$. The change in value has two components using the discrete time 'differentiation rule':

$$\Delta V_t = \phi_t \Delta X_t + X_t \Delta \phi \ .$$

The first changes in portfolio value are due to changes in asset prices. The second one are due to changes in the strategy in a given period, i.e. additional money is injected or withdrawn. If strategies are considered where the second part is zero we speak about self-financing strategies: A self-financing strategy only redistributes total wealth after new prices are known. We restrict ourself **always** to self-financing strategies. If ϕ_t is self-financing, the portfolio value reads

$$V_t = V_0 + \sum_{j=0}^{t} \phi_j \Delta X_j \ .$$

The final portfolio value is equal to the initial value plus the cumulative gains and losses from the price changes of the asset X over time weighted by the investment strategy. If we recall that replication means $V_t = C_t$ in **all** states and at **all** time points, the above equation transforms to - if we can **replicate** -

$$C_t = C_0 + \sum_{j=0}^{t} \phi_j \Delta X_j \ .$$

ϕ_j is the hedging strategy which given the initial option price C_0 generates the random option claims C_t. The **martingale representation theorem** states when such a strategy exists, see below. The notion of an arbitrage strategy carries over from the one-period case. Formally:

Definition 2.4.1. *A self-financing strategy ϕ is an arbitrage strategy, if the portfolio value under this strategy satisfies: $V_0 = 0$, $V_t \geq 0$ for all $t = 1, \ldots, T-1$, $V_T \geq 0$ for all states and $V_T > 0$ for one state.*

Starting with zero initial portfolio value, we face no losses in the future and there is a chance that we end up with a gain.

We consider the evolution of the **information** which is richer than in the one period model. We work with a 3-period model. We can **observe** 8 possible path realizations ω_k, see Figure 2.11:

$$\begin{aligned} \omega_1 &= (u,u,u) \ , \ \omega_2 = (u,u,d) \ , \ \omega_3 = (u,d,u) \ , \ \omega_4 = (u,d,d) \\ \omega_5 &= (d,u,d) \ , \ \omega_6 = (d,d,u) \ , \ \omega_7 = (d,u,d) \ , \ \omega_8 = (d,d,d) \ . \end{aligned}$$

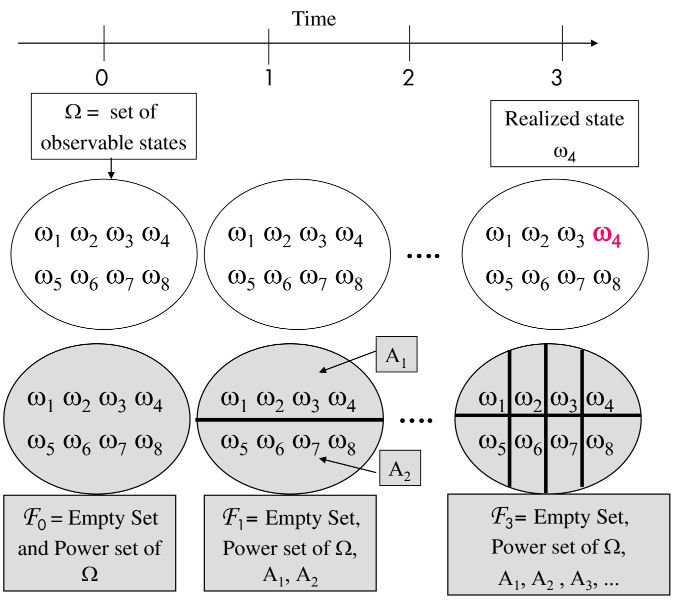


Figure 2.11: Illustration of the information and filtration structure for the three period CRR.

That is **an observable state** is a **path** in the tree. In a k period model there are 2^k paths. We assume that the up and down probabilities are constant over time. Suppose that the first move was up. Then several paths are still possible but other paths are no longer possible, for example the path with 3 down-moves. If in the second step there is a down move, some more paths will become impossible and so on. This shows that besides observable elements also **possible** events are important. The notion of filtration considers the possible event structure and its dynamics over time. If we have 8 observable events, the power set $\mathcal{A} = 2^8$ defines all possible events, i.e. from the empty set to the set of all paths every set is included. $\mathcal{F}_t$ represents the possible information up to time t. This set is an element of the power set. Assume that

$$\mathcal{F}_t \subset \mathcal{F}_{t+1} \ , \mathcal{F}_t \in \mathcal{A} \ \forall t \ .$$

The inclusion $\mathcal{F}_t \subset \mathcal{F}_{t+1}$ means that for increasing time, the information resolution increases. At $t = 0$, $\mathcal{F}_0 = \{\emptyset, \mathcal{A}\}$ - everything is possible, no information resolution so far took place or all information is still random.[7] At $t = 1$, either the price S_0 increased or decreased. We therefore define the sets

$$A_1 = \{\omega_1, \omega_2, \omega_3, \omega_4\} \ , \ A_2 = \{\omega_5, \omega_6, \omega_7, \omega_8\} \ .$$

A_1 (A_2) is the set of all events where the first price move is 'up' ('down'). We set

$$\mathcal{F}_1 = \{\emptyset, \mathcal{A}, A_1, A_2\} \ .$$

This assures that $\mathcal{F}_0 \subset \mathcal{F}_1$. It is clear, how $\mathcal{F}_2$ can is defined. The information sets or filtrations was **generated by the evolution of the asset prices.** Other source such as noise or insider information do not enter the model.

Given a specific date we use the information at this date to forecast future prices using the conditional expectation based on the available information set.

Since for each node on the tree there is a filtration and vice versa, nodes and filtrations are in one-to-one relationship for trees. How does self-financing strategies fit into the filtration concept?

The strategy ψ_t is known in advance at $t-1$ i.e. an element of $\mathcal{F}_{t-1}$. Such strategies are called **predictable** processes. The price processes S_t are not known in advance, i.e. they are observed at t and therefore related to $\mathcal{F}_t$. Such processes are called **adapted**.

The notion of information is fairly intuitive in discrete time models. In continuous time the mathematics becomes intricate.

[7] The inclusion of the empty set guarantees that $\mathcal{F}_0$ has required mathematical structure is powerful, i.e. the information set system is closed under countable intersection and complement set formation.

The First Fundamental Theorem of Finance transforms to the CRR model case. The theorem is based on the notions of **risk neutral probabilities** or **equivalent martingale measures.**

Definition 2.4.2. *Consider a price process under a probability P. A probability Q is equivalent to P, written $Q \sim P$, if they have the same the same impossible sets. A probability $Q \sim P$ is a risk neutral probability if the discounted price process $\tilde{S} := S/N$ is Q-martingale with $N > 0$ the numeraire, i.e.*

$$\tilde{S}_t = E^Q[\tilde{S}_s|\mathcal{F}_t] ,$$

holds for all t and $s \geq t$.

Equivalence means, that $P(\text{State } k) > 0$ for all states implies $Q(\text{State } k) > 0$ for all states and vice versa. The definition of a martingale leads to the following characterizations:

- Martingales are processes with **zero drift.**
- The expected value of a martingale process is **constant**. This follows from the tower law or the law of iterated expectations.

Proposition 2.4.3. *Consider the CRR model and self-financing strategies. There is no arbitrage if and only if there exists a risk neutral probability Q.*

We interprets the martingale property next. Since S_t is known at t, it can be moved inside of the expected value, i.e.

$$E^Q[S_{t+1}/S_t|\mathcal{F}_t] = 1 + r .$$

Taking expectations w.r.t. $\mathcal{F}_0$ on both side we get by the law of iterated expectations:

$$E^Q[S_{t+1}/S_t] = 1 + r \ \Rightarrow \ E^Q[\frac{S_{t+1} - S_t}{S_t}] = r .$$

Therefore, the expected return of the risky asset under the risk neutral probability equals the risk free rate. No arbitrage implies that on average the return process cannot grow in the risk neutral world stronger than the risk free asset. It is useful to introduce the random variable

$$X_{t+1} := S_{t+1}/S_t .$$

This price ratio takes only the values $(1+u)$ or $(1+d)$. Since Q is strictly positive, both values are attained with a positive probability. This implies $d < r < u$, else the equation $E^Q[S_{t+1}/S_t] = 1+r$ does not hold true. The last inequality is the no arbitrage condition of the one period model. If the inequality is violated, arbitrage is possible.
Can we construct the measure Q? The following proposition gives the answer, see Appendix 7.4 for the proof.

Proposition 2.4.4. *Assuming $-1 < d < r < u$. The following statements are equivalent:*

1. *$\tilde{S}_k$ is a Q-martingale.*

2. *The random variables S_{k+1}/S_k are i.i.d. under $Q = (q, 1-q)$ with*

$$
\begin{aligned}
Q[S_{k+1}/S_k = 1+u] &= q = \frac{r-d}{u-d} \\
Q[S_{k+1}/S_k = 1+d] &= 1-q = \frac{u-r}{u-d} .
\end{aligned}
$$

The risk neutral probability is unique, i.e. the CRR market is complete. If the underlying instrument pays a dividend yield $\delta \geq 1$, the risk neutral probability and the no arbitrage condition are:

$$
u > r/\delta > d \ , \ q = \frac{r/\delta - d}{u-d} .
$$

We show how to price an European call option in the CRR model. Such a contract pays at maturity $C(S_T) = \max(S_T - K, 0)$ with K the strike value. The following proposition is proven in Appendix 7.4.

Proposition 2.4.5. *The arbitrage free price of a call option in the n-period CRR model is given by:*

$$
C(S,t) = \sum_{k=0}^{n} \binom{n}{k} q^n (1-q)^{n-k} \max(S_0 u^k d^{n-k} - K, 0) . \tag{2.40}
$$

Separating the two terms in the payoff formula[8], the price of a call option shows the same form as in the one period model. The price is proportional to the underlying value and the present value of the strike. Since this last expression has negative sign it represents a loan. The difference to the one period model are the more complicated factors in front of S and K. They are probabilities. The formula can be read as follows:

The price of a call or put option in the CRR model at a date t with maturity T is given by:

$$
C(S,t) = \sum_{\text{paths}} \text{Path Probability} \times \text{No. of paths} \times \text{Payoff End Node } T \tag{2.41}
$$

where the path probability equals $q^{k_u}(1-q)^{k_d}$ with q the risk neutral probability k_u the number of 'up' moves on the given path from t to T and similarly k_d the number of 'down' moves, 'No. of paths' the number of paths connecting the node at time t with the end node at T.

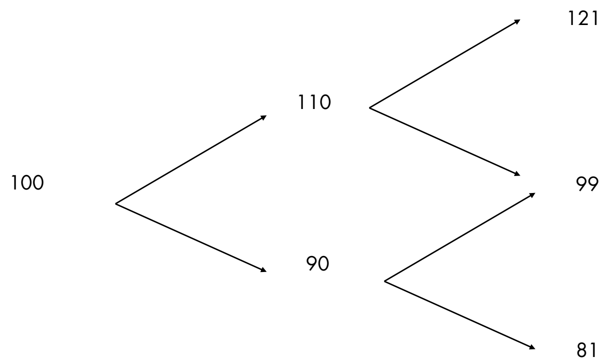


Figure 2.12: Underlying value S in two periods.

2.4.2 Examples

1. T Consider a call option in a two period model with strike 95 in a currency and risk less interest rate of 4% per period. The underlying value dynamics is shown in Figure 2.12.

   The risk neutral probability is $q = \frac{1.04-0.9}{1.1-0.9} = 0.7$. Using (2.41) we get:

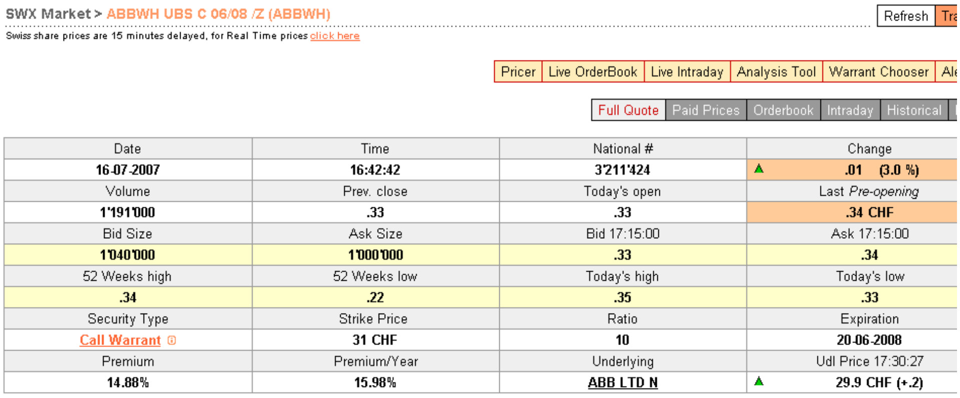
$$C_0 = \frac{1}{(1.04)^2}\left(0.7^2 \times 1 \times 26 + 0.7 * 0.3 \times 2 \times 4 + 0.3^2 \times 1 \times 0\right) = 13.33$$

   where each term is calculated by multiplying (i) the probability of a path with (ii) the number of paths and (iii) the final node option payoff.

2. We compare the accuracy of the binomial CRR model with observed option prices. Consider a call on ABB Ltd. with strike CHF 31 and expiration June, 20 2008. The bid and ask prices where at CHF 0.33 and 0.34 respectively and the actual ABB share price was CHF 29.9, see Figure 2.15. The figures are calculated using (2.41).

   The first step is to calculate the tree u, d, r from the real world data. If R is the annual rate, r the rate on the tree, n the number of periods in the tree and $\tau = T - t$ time to maturity, we have the relationship $r^n = R^\tau$. The number of periods in the CRR model is $n = 11$, time to maturity is $\tau = 0.917$. This implies $r = R^{\tau/n} = 1.00327$. The discount factor is $D = e^{-\tau r/n} = 0.920.$ We need the up and down values. A standard approach is to set $u = e^{\sigma_{1y}\sqrt{\tau/n}}, d = 1/u$, we

[8]Due to the max operator a separation leads to an adjustment of the summation range.


SWX Market > ABBWH UBS C 06/08 /Z (ABBWH)

Swiss share prices are 15 minutes delayed, for Real Time prices click here

Refresh

Pricer | Live OrderBook | Live Intraday | Analysis Tool | Warrant Chooser

Full Quote | Paid Prices | Orderbook | Intraday | Historical

Date	Time	National #	Change
16-07-2007	16:42:42	3'211'424	.01 (3.0 %)
Volume	Prev. close	Today's open	Last *Pre-opening*
1'191'000	.33	.33	.34 CHF
Bid Size	Ask Size	Bid 17:15:00	Ask 17:15:00
1'040'000	1'000'000	.33	.34
52 Weeks high	52 Weeks low	Today's high	Today's low
.34	.22	.35	.33
Security Type	Strike Price	Ratio	Expiration
Call Warrant	31 CHF	10	20-06-2008
Premium	Premium/Year	Underlying	Udl Price 17:30:27
14.88%	15.98%	ABB LTD N	29.9 CHF (+.2)

Figure 2.13: Data for the ABB call. *Source: Swissquote.*

comment about this choice at the end of the example. Using ABB closing data we get a daily volatility of $\sigma_{1d} = 0.0150983$. To obtain an annualized volatility we use the square-root rule, i.e.

$$\sigma_{1y} = \sqrt{\text{days}}\sigma_{1d} = \sqrt{250}\sigma_{1d} = 0.2882$$

where we assumed that there are 250 days in a year. We also need to know the risk less rate for one period. This gives $u = 1.087, d = 0.92$. These values imply for the risk neutral probability $q = 0.499$. The table shows the pricing result.

The sum of the path weights over all end nodes is 1 and the sum of the payoffs over all nodes, i.e. the last column, is CHF 3.41205. Discounting this value back to time zero gives 3.1383. Using the ratio 1 : 10 gives the theoretical price of 0.31 CHF compared to the actual bid-ask prices of $0.33 - 0.34$.

We finally consider the **relationship** between discrete CRR and continuous time modelling. Consider a continuous time model for the risky asset where the mean and variance of the asset ratio S_{t+dt}/S_t are given by

$$\text{Mean}(S_{t+dt}/S_t) = e^{rdt} \ , \ \text{Variance}(S_{t+dt}/S_t) = e^{2rdt}(e^{\sigma^2 dt} - 1)$$

where r is a risk free interest rate and σ is the volatility of the continuous time price process of the risky asset. In the Black and Scholes, which turns out to be

Node	$S_0u^{k-10}d^k$	$max(S_T - K, 0)$	No. of Paths	$q^{k-10}(1-q)^k$	S.P.	Sum Payoff
11	74.541	43.541	1	0.0004	0.0004	0.020
10	63.114	32.114	11	0.0004	0.0052	0.168
9	53.440	22.440	55	0.0004	0.0264	0.593
8	45.248	14.248	165	0.0004	0.0796	1.134
7	38.312	7.312	330	0.0004	0.1600	1.170
6	32.440	1.440	462	0.0004	0.2250	0.324
5	27.467	-	462	0.0004	0.2260	0.000
4	23.257	-	330	0.0004	0.1622	0.000
3	19.692	-	165	0.0004	0.0814	0.000
2	16.673	-	55	0.0004	0.0272	0.000
1	14.118	-	11	0.0004	0.0054	0.000
0	11.954	-	1	0.0005	0.0005	0.000
				Sum	1	3.41205

Table 2.4: Valuation of the call option in the 11 period model with ABB as underlying value. 'S.P.' means 'sum of path weights'.

a continuous time limit of the CRR model, the risky asset dynamics will satisfy exactly this dynamic properties under the risk neutral measure. Contrary, in the one period model the mean and variance of the same asset ratio are[9]

$$\text{Mean}(S_{t+1}/S_t) = qu + (1-q)d$$

and

$$\text{Variance}(S_{t+1}/S_t) = qu^2 + (1-q)d^2 - (qu + (1-q)d)^2 = q(1-q)(u-d)^2 \ .$$

Equating the two moments in both models we get:

$$qu + (1-q)d = e^{rdt} \ , \ q(1-q)(u-d)^2 = e^{2rdt}(e^{\sigma^2 dt} - 1) \ .$$

This implies the risk neutral probability formula $q = \frac{e^{rdt}-d}{u-d}$ from expectation matching. Making the **symmetric choice** $u = 1/d$ from the variance matching condition a complicated expression for u follows. Using a Taylor approximation by assuming that dt is small one gets

$$u = 1/d = 1 + \sigma\sqrt{dt} + \frac{\sigma}{2}dt + \ \dots \ .$$

The first terms agree with the power series expansion of $u = e^{\sigma\sqrt{dt}}$ - this justifies the approach in the above pricing of the ABB call option.

[9] $\text{var}(X) = E[X^2] - E^2[X]$.

3. We consider **Delta hedging**, first in a one period and then in the CRR model. We consider a portfolio of N call options $C(S)$ with underlying S. The portfolio value V is

$$V = N * C(S) \tag{2.42}$$

with N the number of option contracts. We assume a ratio of 1, i.e. one option gives the right to buy one unit of the underlying value. To Delta hedge the portfolio we search for a number n_S of underlying S such that the augmented hedge portfolio V^h

$$V^h = n_S + NC(S) \tag{2.43}$$

is Delta risk free: $\Delta(V^h) = 0$. Taking the derivative, the Delta reads

$$\Delta(V^h) = n_S \Delta S + N * \Delta C(S) = n_S + N * \Delta C(S) \ .$$

Delta neutrality implies:

$$\Delta(V^h) = 0 \iff n_S + N\Delta(C(S)) = 0 \iff n_S = -N\Delta(C(S)) \ . \tag{2.44}$$

We next consider Delta in a multi period CRR model. That is we choose a strategy Δ_t such that at each date $t-1$ fixing the Delta at this date and the investment in the risk less asset delivers a portfolio value which is equal to the claim or option value at the next step t in all possible states. Choosing a $\Delta_t^{u/d}$ means to buy or sell an amount of the underlying stock at time t, where we have to consider the last period of the stock. The trader is Delta neutral if

- at $t-1$: A Delta Δ_{t-1} and an amount ϕ_{t-1} of the risk less asset are chosen such that
- at t:

$$\Delta_{t-1} \times S_t + \phi_{t-1} \times \text{risk less return} = \text{Payoff in } t.$$

At time $t = 2$, using backward induction, we calculate $\phi_1, \Delta_1^{u/d}$. Consider a call option in a two period model with strike 95 in a currency and risk less interest rate of 4% per period. The underlying value dynamics is shown in Figure 2.12. In the upper area of the tree in Figure 2.12, we have:

$$\begin{aligned} 26 &= \Delta_1^u 121 + \phi_1^u 1.04^2 \\ 4 &= \Delta_1^u 99 + \phi_1^u 1.04^2 \ . \end{aligned}$$

Solving gives $\Delta_1^u = 1, \phi_1^u = -\frac{95}{1.04^2}$. If we are in the lower area, that is the stock went down in the first period, we have:

$$\begin{aligned} 4 &= \Delta_1^d 121 + \phi_1^d 1.04^2 \\ 0 &= \Delta_1^d 99 + \phi_1^d 1.04^2 \ . \end{aligned}$$

The solution of this system is $\Delta_1^d = 2/9, \phi_1^d = -\frac{18}{1.04^2}$. At time $t = 1$, we have to solve:

$$\begin{aligned}\Delta_0 110 + \phi_0 1.04 &= \Delta_1^d 110 + \phi_1^d 1.04 \\ \Delta_0 90 + \phi_0 1.04 &= \Delta_1^d 90 + \phi_1^d 1.04 \ .\end{aligned}$$

The solution of this equations gives $\Delta_0 = \frac{83}{104}, \phi_0 = -\frac{71.9}{1.04^2}$. We show that this Delta hedging strategy is self-financing and replicating. We assume that the following path is realized: In the first period the risky assets moves up and it falls in the second period to 99. Using the calculated Delta we have at time 0:

$$V_0 = 83/104 * 100 - 71.9/104^2 = 13.33 = C_0 \ .$$

i.e. we can finance the fair call price with the hedge. At time 1 we have:

$$V_1^u = 83/104 * 110 - 71.9/104 = 18.65 \ .$$

This portfolio value has to be reallocated for the next period using the Delta of 1 and the risk less strategy of $-95/1.04^2$. The price of the new portfolio is

$$V_1 = 1 * 110 - 95/104 = 18.65 \ .$$

Hence the reallocation is self-financing. At time 2 the new portfolio has the value

$$V_2 = 1 * 99 - 95 = 4 \ ,$$

i.e. the payoff of the call option at maturity in all states is replicated.

4. We consider a **barrier option**, more precisely with a **down& out call (DUC) European option** in a 3 period model. Barrier options are path dependent which follows from the payoff definition. A DUC is parameterized by a strike K like an ordinary or vanilla call option and a barrier B. The barrier is lower than the strike - the 'down' - and the option is worthless if during the life time of the option the barrier is hit at least once - the 'out'. To price such an option we assume $B = K$, i.e. a **knock-out option**. Figure 2.14 shows the necessary data.

   We assume $u = 1.25, d = 0.8, r = 1.1$. This implies the stock price path in the left panel of Figure 2.14. The right panel shows the intrinsic values of the option. It follows that for different paths different option values follow - there is a path dependency. The dashed path hits the barrier: The option becomes worthless. If we consider the dotted paths we end up at the same final node but we never hit the barrier which give the intrinsic value of 180. Therefore

   $$\text{up x down} \neq \text{down x up}$$

   for path dependent options. To value this option we neglect all paths which lead to a positive final payoff which once hit the barrier. We have for the top end node

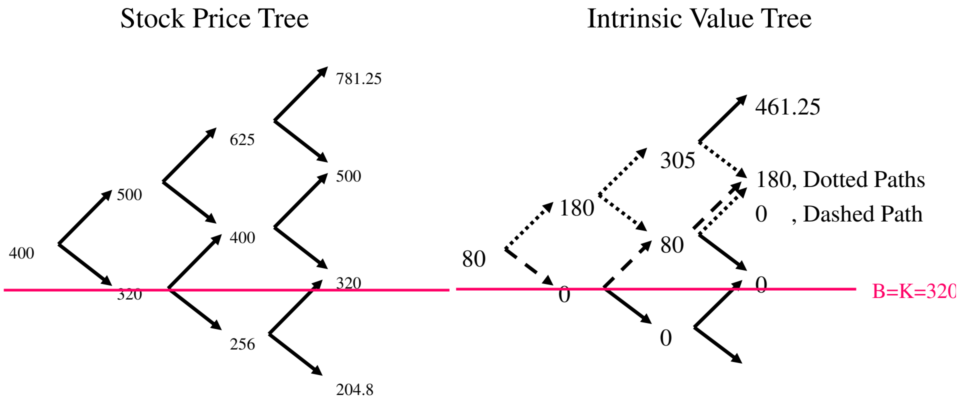


Figure 2.14: Knock out barrier option.

481.25 a single path which never breached the barrier. For node 180 there are 3 paths but only 2 of them are never knocked-out. All other nodes have zero terminal payoff. The risk neutral probability is $q = 0.444$. The option price is therefore:

$$\begin{aligned} C_0^{\text{DUC}} &= \frac{1}{(1.1)^3} \\ &\times \left(0.444^3 \times 1 \times 461.25 + 0.444^2 \times 0.555 \times 2 \times 180\right) = 59.9 \ . \end{aligned}$$

The price of the vanilla call is 74.7. This is a general fact that barrier options are more risky and hence less expensive than their vanilla counter parts.

5. **American Put.** Consider a binomial tree with n time steps. Let $S_k(j)$ be the price of the underlying value at time k in node j. The intrinsic value of a vanilla put option is $K - S_k(j)$. Since it might be optimal to early exercise the option the investor faces at each time and in each node the decision problem whether to continue or to exercise. If $P(k, j)$ is the arbitrage free put option price at time k in state j, the decision problem reads:

$$P(k, j) = \max(E^Q(P(k+1)D(k, k+1)), X - S_k(j)) \ .$$

Since we consider a recombining tree,

$$E^Q(P(k+1)D(k, k+1)) = \frac{1}{1+r}\left(qP(k+1, j+1) - P(k+1, j)\right)$$

holds for a constant discount factor. Why is this not equal to

$$E^Q(P(k+1)D(k,k+1)) = \frac{1}{1+r}(qP(k+1,j+1) - P(k+1,\mathbf{j\text{-}1}))\ ?$$

From k to $k+1$ time step the node j moves to the nodes j and $j+1$ in $k+1$. This expected value is the **continuation value.** This defines the backward induction algorithm to price an American put - at each node the investor has to decide whether to continue or to realize. Consider a stock, the figures of this and the following example are from Kwok (2011), with strike and initial price S_0 50 in a currency, a 10 percent risk free rate, 40 percent volatility and 5m maturity. Consider a 5 period tree, i.e. $\Delta t = 0.0833$ is the length of one period equal to one month. Then $u = e^{\sigma\sqrt{\Delta t}} = 1.1224, d = 1/u = 0.8909, R = e^{r\Delta t} = 1.0084, q = 0.5073$ follow as input parameters.

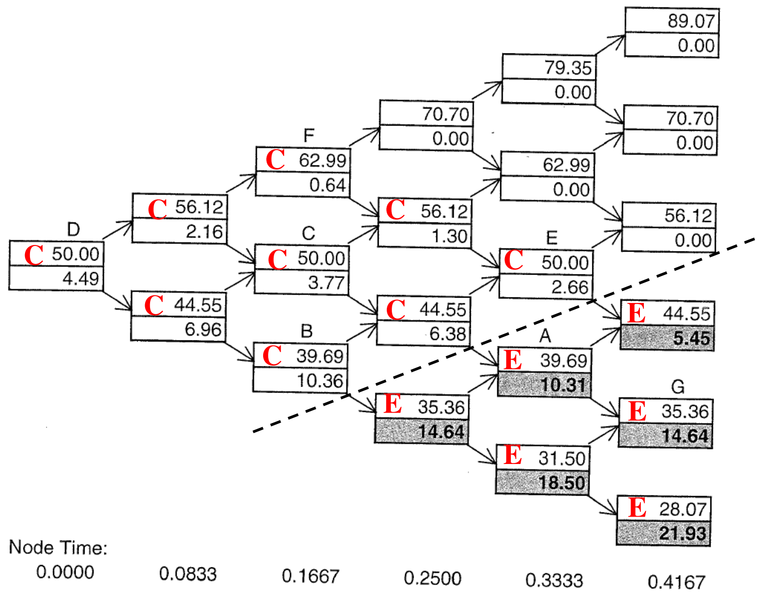


Figure 2.15: American put. At each node the value of the underlying (upper number) and the value of the option (lower value) are shown. The letter 'E' indicates that execution is optimal at the node and 'C' that continuation is optimal. The dashed line shows that the boundary region between the two decision. *Source: Y.K. Kwok, 2011, adapted Figure..*

The stock price $S_k(j)$ is given by j up-moves and $k-j$ down-moves, i.e. $S_k(j) = S_0 u^j d^{k-j}$. The final option prices are simply $\max(K - S_T, 0)$. The option value without considering early exercise is say at node E equal to

$$e^{-0.1\times 0.0833}(0.5073 \times 0 + 0.947 \times 5.45) = 2.66\ .$$

Since early exercise at this node gives zero, it is optimal to wait in node E. The figure shows the critical boundary between exercising early and continuation. A variant of this product is a **callable American put**, i.e. a feature that entitles the issuer to buy back the American option at any time at a predetermined call price. Upon call, the holder can choose either to exercise the put or receive the put price as cash. Let X be the price if the put is called. Then the simple American put recursion becomes

$$P(k,j) = \min\left(\max(E^Q(P(k+1)D(k,k+1)), K - S_k(j)), \max(X, K - S_k(j))\right) .$$

The first max-term represents the optimal strategy of the holder, given no call of the option by the issuer, the second is the payoff if the product is called where the holder receives either the cash amount X or exercises the option. The issuer chooses to call or not in order to maximize the option value with reference to the possible actions of the holder. The value of the callable call is given by taking the minimum value of the above two terms. The recursion is equivalent to

$$P(k,j) = \max\left(\min(E^Q(P(k+1)D(k,k+1)), X), K - S_k(j)\right) .$$

The issuer calls the option when the continuation value is above the call price X. When the option is either called or not called, the holder can always choose to exercise to receive $K - S_k(j)$ if the exercise payoff has a higher value.

6. So far the examples were equity based. We turn to an interest rate example. Consider the spot LIBOR rate $L(t,T)$ at time t for maturity T defined by

$$L(t,T) = \frac{1 - p(t,T)}{(T-t)p(t,T)} . \tag{2.45}$$

We assume a constant grid of dates $T_j, j = 1, \ldots, n$, i.e. $T_j - T_{j-1} = \delta$ for all j. For each date exists a zero $p(t,T_j)$. The forward LIBOR rate $F(t, i-1, i)$ at time t with expiry T_{i-1} and maturity T_i is given by

$$F(t, i-1, i) = \frac{1}{\delta}\left(\frac{p(t,T_{i-1})}{p(t,T_i)} - 1\right) . \tag{2.46}$$

The definition can be written as

$$1 + \delta F(t, i-1, i) = \frac{p(t,T_{i-1})}{p(t,T_i)} ,$$

which is the no arbitrage relationship between forward and bond prices. The LIBOR market model models the family of rates $F(t, i-1, i)$ for a spanning number of dates i.

We show how no arbitrage pricing works in this two period model. We follow Neftci (2008).

- Model with dates T_0, T_1, T_2.
- Two zero bonds $p(T_0, T_i)$ where one matures at time T_1 and the other one at time T_2.They are assumed to be default free.
- A money market account B that pays in-arrears the discrete time simple LIBOR spot rate $L(T_i, T_{i+1})$ observed at time T_i. Payoff of the account at time T_2 is the random variable

$$B(T_2) = (1 + \delta L(T_0, T_1))(1 + \delta L(T_1, T_2)) \ .$$

- A FRA contracted at time T_0, settled at time T_2. The buyer receives/pays the difference between the fixed forward rate $F(0, T_1, T_2)$ and the floating rate $L(T_1, T_2)$ at time T_2. The final payoff with notional 1 reads for the buyer

$$\delta(L(T_1, T_2) - (F(0, T_1, T_2)) \ .$$

- The underlying random LIBOR rates follow a binomial model with u, d moves.

In this market the payoff matrix $\mathbb{P}_{T_2} =: \mathbb{P}_2$ reads, see (2.47):

$$\mathbb{P}_2' = \begin{pmatrix} B_2^{uu} & B_2^{ud} & B_2^{du} & B_2^{dd} \\ 1 & 1 & 1 & 1 \\ p_2^{uu} & p_2^{ud} & p_2^{du} & p_2^{dd} \\ \delta(F_0 - L_1^u) & \delta(F_0 - L_1^u) & \delta(F_0 - L_1^d) & \delta(F_0 - L_1^d) \end{pmatrix} \tag{2.47}$$

for the money market account, the T_1-zero, the T_2-zero and the FRA. We simplified the notation as follow: $B_2^{uu} = B(T_2)^{uu}$, $p_2^{uu} = p(T_0, T_2)^{uu}$, $F_0 = F(0, T_1, T_2)$ and $L_1^u = L(T_1, T_2)^u$. If there is no arbitrage, the First Fundamental Theorem of Finance states there exists a state price vector $\psi = (\psi^{uu}, \psi^{ud}, \psi^{du}, \psi^{dd})$ with strictly positive components such that $\mathbb{P}_2'\psi = S_0$. Explicitly we have the linear pricing relationship:

$$\begin{pmatrix} 1 \\ p(T_0, T_1) \\ p(T_0, T_2) \\ 0 \end{pmatrix} = \begin{pmatrix} B_2^{uu} & B_2^{ud} & B_2^{du} & B_2^{dd} \\ 1 & 1 & 1 & 1 \\ p_2^{uu} & p_2^{ud} & p_2^{du} & p_2^{dd} \\ \delta(F_0 - L_1^u) & \delta(F_0 - L_1^u) & \delta(F_0 - L_1^d) & \delta(F_0 - L_1^d) \end{pmatrix} \begin{pmatrix} \psi^{uu} \\ \psi^{ud} \\ \psi^{du} \\ \psi^{dd} \end{pmatrix} .$$

If we use the risk neutral measure associated with the money market account, the risk neutral probabilities are obtained from the first row:

$$1 = B_2^{uu}\psi^{uu} + B_2^{ud}\psi^{ud} + B_2^{du}\psi^{du} + B_2^{dd}\psi^{dd} \ .$$

Since the $\psi' s$ are not negative, $q^{uu} := B_2^{uu}\psi^{uu}$ and the other three q's are indeed probabilities. They defines a probability Q. If we use the paths $\omega_1 = uu, \omega_2 = ud$ and so on we can write

$$1 = \sum_{k=1}^{4} q(k) = \sum_{k=1}^{4} B_2(k)\psi(k) \ .$$

Using the risk neutral probabilities, the last row in the linear pricing relationship for the FRA reads:

$$0 = \delta \sum_k (F_0 - L_1^\sharp)\psi(k) = \delta \sum_k (F_0 - L_1)\frac{q(k)}{B_2(k)} = \delta E^Q \left[\frac{F_0 - L_1}{B_2}\right]$$

where $L_1^\sharp$ equals L_1^u in the states ω_1, ω_2 and down in the other two states. Solving this last equation w.r.t. the forward rate we get

$$F_0 = \frac{1}{E^Q\left[\frac{1}{B_2}\right]} E^Q \left[\frac{L_1}{B_2}\right] .$$

Since B_2 is a random variable, the last expression cannot be simplified to

$$F_0 = E^Q [L_1] .$$

Hence, under the risk neutral measure associated to the money market account the forward rate is not equal to the estimated future LIBOR rate. This bias is due to convexity. Generalizing, the forward rate F_t is **not** a martingale under the above specified risk neutral probability Q. Therefore, if we model the dynamics of the forward rates F_t we have to model a drift and to calibrate this drift. This makes the measure Q an inconvenient pricing tool. The **forward measure** is a more adequate measure. This measure is defined for each zero bond and hence different for different bond maturities. We choose the T_1-bond. The second row of the pricing equation reads

$$p(0, T_1) = \sum_k \psi(k) .$$

Dividing by the bond price this defines the $Q^{T_1} =: Q^1$-forward measure probabilities, i.e.

$$1 = \sum_k \frac{\psi(k)}{p(0, T_1)} =: \sum_k q^1(k) .$$

The general theory implies that any asset A with payout only at time T_1 can be priced by

$$A_0 = p(0, T_1) E^{Q_1}[A_T] ,$$

i.e. the ratio $Z_t = \frac{A_t}{p(t,T)}$ is a martingale under the forward measure Q_1. Using this measure we arrive at

$$F_0 = E^{Q_1}[L_1] .$$

To obtain this divide the fourth row of the pricing equation by the T_1-zero and rearrange. Under the forward measure the forward LIBOR rate is a martingale.

2.4.3 Measure Change and Hedging

We encountered the need to change the measure in pricing derivatives. To understand the technique we start with the **change of measure technique** in a two period model setup. This is an extended version of Baxter and Renyj (2003). p_1 is the probability for up in the first time step, p_2 for up in the second one, given the first move was up and p_3 the up probability in the second step, given the first move was down. The filtration is generated by the price dynamics:

$$\mathcal{F}_0 = \{\emptyset, \Omega\} \subset \mathcal{F}_1 = \{\emptyset, \Omega, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\} \subset \mathcal{F}_2$$

where $\mathcal{F}_2$ equals the power set of the four elements ω_i. Each ω represents a **possible price paths**in the two period model. There are 4 paths. Instead of considering the up- and down-probabilities, we switch to path's probabilities since derivative prices are expected values of final payoff values times the number of paths and the path probabilities. We have:

Realizations	Path	Probability P and Π
up-up	ω_1	$p_1 p_2 =: \pi_1$
up-down	ω_2	$p_1(1-p_2) =: \pi_2$
down-up	ω_3	$(1-p_1)p_3 =: \pi_3$
down-down	ω_4	$(1-p_1)(1-p_3) =: \pi_4$

We defined a path probability measure $\Pi = (\pi_1, \ldots, \pi_4)$ out of the price process probability $P = (p_1, \ldots, p_3)$. The sum over all path probabilities equals one. If $0 < p_i < 1$ for all i, given the p's we can obtain the π's and vice versa: P and Π are **equivalent**. Suppose that we deform the P-probabilities to a new probability $Q = (q_1, \ldots, q_3)$. If $0 < q_i < 1$, P and Q are equivalent and the Q's generate new path probabilities $\tilde{\Pi}$. These new path probabilities are equivalent to the former ones since the equivalence of probabilities is a transitive relation.

- For each path i, the **density** $Z_i = \frac{\tilde{\pi}_i}{\pi_i} =: \frac{dQ}{dP}$ (Radon-Nikodym derivative) measures the difference between the two path probabilities.

- This derivative allows us to switch from P to Q as follows. Knowing P, we know π_i. The Radon-Nikodym derivative then gives the ratio $\tilde{\pi}_i/\pi_i$. Hence, we know $\tilde{\pi}_i$ which allows us to calculate Q.

- Suppose that $\tilde{X}(T)$ is a discounted contingent claim. The no arbitrage price under $\tilde{\Pi}$ satisfies

$$\tilde{X}(0) = E^{\tilde{\Pi}}[\tilde{X}(T)] = \sum_i \tilde{\pi}_i \tilde{X}_i(T) = \sum_i \pi_i \frac{\tilde{\pi}_i}{\pi_i} \tilde{X}_i(T) = E^{\Pi}[Z\tilde{X}(T)] \ .$$

- The density serves to calculate the new probabilities as follow:

$$P \mapsto \Pi \text{ and } Z \mapsto \tilde{\Pi} \mapsto Q \ .$$

So far we studied the density Z at maturity. But $Z := Z_t$ is a stochastic process which defines at each date the 'stretching factor' between two equivalent probabilities. If the probabilities q_j are all equal to $\frac{1}{2}$, then all paths $\tilde{\pi}_i$ have the same probability under Q, i.e. $\tilde{\pi}_i = \frac{1}{4}$. If the price process probabilities p_j are all equal to $\frac{1}{4}$ we get for the path probabilities π_i under P

$$\pi_1 = \frac{1}{16} \;,\; \pi_2 = \frac{3}{16} \;,\; \pi_3 = \frac{3}{16} \;,\; \pi_4 = \frac{9}{16} \;.$$

Using this we calculate the density process Z_2 for the path probabilities Π and $\tilde{\Pi}$ at time $t = 2$:

$$Z_2(\omega_i) := \frac{dQ_2(\omega_i)}{dP_2(\omega_i)} = \frac{\tilde{\pi}_i|_{\mathcal{F}_2}}{\pi_i|_{\mathcal{F}_2}} = \begin{cases} 4, & \omega_1, \\ \frac{4}{3}, & \omega_2, \\ \frac{4}{3}, & \omega_3, \\ \frac{16}{36}, & \omega_4 \end{cases} .$$

The figures are calculated as follow. For the state ω_1 we have (see Figure 2.16):

$$Z_2(\omega_1) = \frac{\tilde{\pi}_1|_{\mathcal{F}_2}}{\pi_1|_{\mathcal{F}_2}} = \frac{q_1 q_2}{p_1 p_2} = \frac{1/2 \times 1/2}{1/4 \times 1/4} = 4 \;.$$

At time $t = 1$ we get:

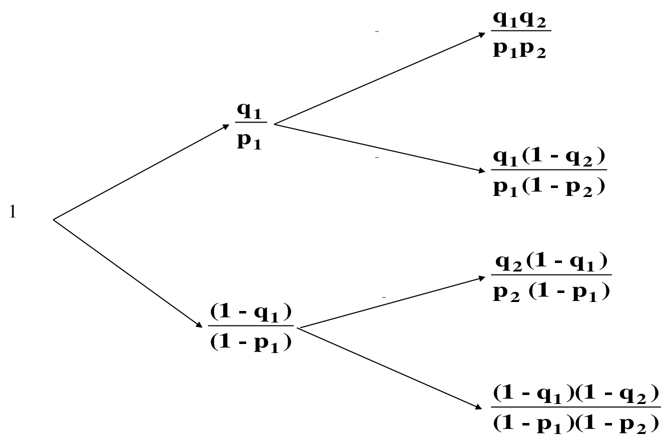


Figure 2.16: The process Z_t in the example.

$$Z_1(\omega_i) = \frac{\tilde{\pi}_i|_{\mathcal{F}_1}}{\pi_i|_{\mathcal{F}_1}} = \begin{cases} 2, & \omega_1, \omega_2 \\ \frac{4}{6}, & \omega_3, \omega_4 \;. \end{cases}$$

We obtain $E_P[Z_2] = 1$. The process Z is a P-martingale. We prove $E^P[Z_2|\mathcal{F}_1] = Z_1$. For the states ω_1, ω_2:

$$E^P[Z_2|\mathcal{F}_1(\omega_1, \omega_2)] = \frac{1}{4} * 4 + \frac{3}{4} * \frac{4}{3} = 2 = Z_1(\omega_1, \omega_2) \ .$$

For the other states:

$$E^P[Z_2|\mathcal{F}_1(\omega_3, \omega_4)] = \frac{1}{4} * \frac{4}{3} + \frac{3}{4} * \frac{16}{36} = 2/3 = Z_1(\omega_3, \omega_4) \ .$$

This proves $E^P[Z_2|\mathcal{F}_1] = Z_1$. The other martingale properties are proven in the same way. How does the 'pricing relation of the last period' $E^Q[\tilde{X}(T)] = E^P[Z\tilde{X}(T)]$ generalizes to the case where we have a process, i.e. how can we relate $E^Q[\tilde{X}_t|\mathcal{F}_s]$ and $E^P[\tilde{X}_t|\mathcal{F}_s]$. One could guess:

$$'E^Q[\tilde{X}_t|\mathcal{F}_s] = E^P[Z_t\tilde{X}_t|\mathcal{F}_s]' \ .$$

This is wrong. The reason is that Z_t describes the scaling from time 0 up to time t. But we calculate the expectations starting in time s given by the filtration. We have to correct for the scaling from 0 to s. This then leaves us with the required amount of change of measure from time s up to time t. To achieve this we simply consider Z_t/Z_s. Consider the two period example for $Z_2(\omega_1)/Z_1(\omega_1, \omega_2)$. Then

$$Z_2(\omega_1)/Z_1(\omega_1, \omega_2) = \frac{q_1 q_2}{p_1 p_2} \frac{p_1}{q_1} = \frac{q_2}{p_2}$$

gives the correct weighting of the measure change from time 1 up to time 2. Summarizing, Bayes' formula follows:

$$E^Q[\tilde{X}_t|\mathcal{F}_s] = Z^{-s} E^P[Z_t\tilde{X}_t|\mathcal{F}_s] \tag{2.48}$$

is the correct formula. Since Z^{-s} is known at time s, it can be taken out of the expected value. The results of the example hold in general. We summarize.

Proposition 2.4.6. *The process*

$$Z_t(\omega) := E_P[Z_T|\ \mathcal{F}_t] \tag{2.49}$$

is a P-martingale and $E_P[Z_T] = 1$. For any random variable $X_t \in \mathcal{F}_t$

$$E_Q[X_t] = E_P[Z_t X_t] \tag{2.50}$$

and Bayes' formula holds:

$$E_Q[X_t|\mathcal{F}_s] = E_P[\frac{X_t Z_t}{Z_s}|\mathcal{F}_s] = \frac{1}{Z_s} E_P[X_t Z_t|\mathcal{F}_s] \ , \ 0 \leq s \leq t \ . \tag{2.51}$$

When can we hedge an option dynamically? The answer is given by the **martingale representation theorem**. Replication means that we can achieve the value of an option X_T at maturity by a selecting a strategy ψ with portfolio value V_t over time such that:

- The portfolio value equals the option value at any date.
- The changes in portfolio value are due to changes in the asset prices only (self-financing strategy).

Formally, when does a self-financing ψ exist such that

$$X_T = V_T = V_0 + \sum_j \psi_j \Delta S_j \ ?$$

In order that the change in prices on the tree is able to match the option, the information structure generated by the price process and the option random variable have to be related: At each time t given the filtration generated by the price process we need to know what happens to the option random variable. Else there is no possibility to adjust the price process in order to generate the desired option values. This means that the option prices needs to be adapted to the filtration generated by the price process. The martingale representation theorem then states that there exists a **Brownian motion** W_t process which has this spanning property: For each random variable which is adapted to the Brownian motion information structure[10] there exists a predictable strategy ψ such that X is spanned by the Brownian motion

$$X = E(X) + \int_0^\infty \psi_s dW_s \ .$$

This shows that formally Brownian motion acts as a 'basis vector' in continuous time. Although the theorem states the existence of a ψ it does not tell us how to find the hedge. What is a Brownian motion? We return to this point later on. For the moment we summarize that specific random variables (martingales) can be represented linearly as sums of a specific stochastic process (Brownian motion).

To gain some more insight we consider a discrete time model on a tree. Let R_t and S_t be two binomial processes. Both are martingales w.r.t. to the same R-filtration. Then by the martingale representation theorem there exists a predictable process ψ such that

$$S_t = S_0 + \sum_{s=1}^{t} \psi_s \Delta R_s \ . \tag{2.52}$$

Why is this useful for hedging? Let X be the option payoff at maturity. Then the process $V_t = E[X|\mathcal{F}_t]$ is a Q-martingale[11] w.r.t. the filtration $\mathcal{F}$ generated by the discounted price

[10] X needs also to be square integrable.

[11] Square-integrability follows from Jensen's inequality:

$$E[V_t^2] = E[E([X|\mathcal{F}_t])^2] \leq E[X^2] < \infty \ .$$

process S/B which is also a Q-martingale. By the martingale representation theorem there exists a ψ such that

$$V_t = V_0 + \sum_{s=1}^{t} \psi_s \Delta S_s / B_s \ . \tag{2.53}$$

Therefore

$$V_T = X = V_0 + \sum_{s=1}^{T} \psi_s \Delta S_s / B_s \tag{2.54}$$

shows that ψ_s is a perfect hedge. We define the strategy (the theorem does not tells us how to find the hedge):

- We hold ψ_t units of the risky asset S_t at time t.
- We hold $\phi_t = X_t/B_t - \psi S_t/B_t$ units of the risk less asset at time t.

This strategy replicates the option and is self-financing. To see this insert the strategies in $V_t = \phi_t B_t + \psi_t S_t$.

2.4.4 Continuous Limit of CRR

The CRR model converges under an appropriate limit procedure to the continuous time and continuous state space model of **Black and Scholes**. The limit is twofold: Discrete time spacing and discrete states become continuous.

We fix time T of the continuous model. We to make sure in the limit procedure that (i) the value of one dollar in $[0, T]$ is the same in the CRR model and in the Black and Scholes model and (ii) that the CRR-price process S_t converges towards a continuous price process which is log-normally distributed (this is the assumed distribution in Black and Scholes).

We first consider (i) and divide $[0, T]$ in m equidistant subintervals, i.e. a binomial model in m periods. We have $(B_k, S_k)_{k=0,\dots,m}$with $B_m = B_T$, $S_m = S_T$. How do we adjust the parameters r_m, u_m, d_m of this model such that they can be compared with the parameters of the CRR model? Let $R = \ln(1 + r)$ be the risk free interest, the instantaneous interest rate with respect to one time unit $[0, 1]$. Setting $r_m := \frac{RT}{m}$ and $B_m = (1 + r_m)^m$ we have in the limit

$$\lim_{m\to\infty} (1 + r_m)^m = \lim_{m\to\infty} \left(1 + \frac{RT}{m}\right)^m = e^{RT} = (1 + r)^T \ ,$$

i.e. B_m converges towards the same value as in the T-maturity continuous model.

We consider (ii) and define relations between r_m, u_m, d_m:

$$\begin{aligned} \ln(1+u_m) &= \ln(1+r_m) + \sigma\sqrt{\frac{T}{m}} \\ \ln(1+d_m) &= \ln(1+r_m) - \sigma\sqrt{\frac{T}{m}} \end{aligned}$$

with a constant $\sigma > 0$, which is independent of m and we set as in the T-model

$$\hat{p}_m := \frac{r_m - d_m}{u_m - d_m}.$$

The above parametrization implies $u/d = 1$. The definitions guarantee that $\log X_i$ is normally distributed with a mean and variance that reduces to the same one as in the assumed risky asset dynamics of the Black and Scholes model, i.e. a geometric Brownian motion. Formally, we have:

Proposition 2.4.7. *Let r_m, u_m, d_m defined as above. Then*

$$\lim_{m\to\infty} \ln\left(\frac{\tilde{S}_T}{S_0}\right) \sim \mathcal{N}(-\frac{\sigma^2 T}{2}, \sigma^2 T).$$

That is, the random variable $\ln\frac{\tilde{S}_T}{S_0}$ converges for $m \to \infty$ in probability to a normally distributed random variable with mean $-\frac{\sigma^2 T}{2}$ and variance $\sigma^2 T$.

The proof is given in Appendix 7.4. We apply this result to price call and put options in Black and Scholes model.

Proposition 2.4.8 (Black-Scholes Formula)**.** *Let u_m, d_m, r_m, $\hat{p}_m$ be given as above. The prices of European call and put options in the Black and Scholes model are given by:*

$$\begin{aligned} \lim_{m\to\infty} C_0^{(m)} &= S_0\Phi(d_1) - Ke^{-RT}\Phi(d_2) \qquad (2.55) \\ \lim_{m\to\infty} P_0^{(m)} &= Ke^{-RT}\Phi(-d_2) - S_0\Phi(-d_1) \end{aligned}$$

with d_1, d_2 given by:

$$\begin{aligned} d_1 &= \frac{\ln(S_0/K) + RT}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2} \\ d_2 &= d_1 - \sigma\sqrt{T} = \frac{\ln(S_0/K) + RT}{\sigma\sqrt{T}} - \frac{\sigma\sqrt{T}}{2} \end{aligned}$$

The proof is given in Appendix 7.4.

2.5 No Arbitrage in Continuous Time: Black and Scholes

2.5.1 Interpretation of the Formula

We have shown how the continuous time, continuous state model of Black and Scholes follows as a limiting model from the CRR. There are two reasons to consider this limit model in more details. First, the ideas of hedging are more neatly explained using integrals than clumsy summation formulae. Second, a lot of finance theory is written in continuous time. In the last years one observes a switch to discrete time modelling.

Consider the call option formula in (2.55). The formula looks complicated but we show that some basic thoughts allow us to understand it. Consider an out-of-the-money call option with price

$$C(S = 90, K = 100, r = 2\%, \sigma = 20\%, t = 0, T = 6 \text{ m}) = 1.99 \sim 2 \ .$$

This positive price cannot be due do interest rates only since investing CHF 90 for 6m gives $90 \cdot e^{r(T-t)} = 90.905$ CHF which is much less than 100. The reason for a price of 2 is due to the fact the underlying is a random variable which has a potential to grow above the strike value in the next 6m. To make this transparent, we assume that the return of the underlying is normally distributed (one of the Black and Scholes assumptions), i.e.

$$\ln(S_T/S_t) \sim N(\mu, \sigma_T)$$

This implies that the stock price at T is log-normal distributed:

$$S_T \sim S_t LN(\mu, \sigma_T) = LN(\ln(S_t) + \mu, \sigma_T) \ .$$

Since we know the distribution, we can price the call using the no arbitrage principle. The price is given

$$C(S, K, r, \sigma, t, T) = E^Q[\max(S_T - K, 0)] \ . \tag{2.56}$$

How do we find the risk neutral probability? No arbitrage implies that the discounted price process S is a martingale with the risk free interest rate as numeraire. But this means that the expected value S has to grow like the risk less asset - else the drifts are not the same. But if the drifts are not the same, their ratio - S/risk less asset - cannot be driftless. Summarizing, we must have at T

$$E[S_T] = S_t \exp(r(T - t)) \ . \tag{2.57}$$

But the expectation of log normal distributed random variable is given by

$$E[S_T] = S_t \exp(\mu + \sigma_T^2/2) \ . \tag{2.58}$$

Equations (2.57) and (2.58) imply:

$$\mu + \sigma_T^2/2 = r(T - t) \quad \Rightarrow \quad \mu = r(T - t) - \sigma_T^2/2 \tag{2.59}$$

The volatility σ_T from t to maturity is determined from the annual maturity by the square-root rule:

$$\sigma_T = \sigma\sqrt{T-t}$$

with σ the annualized volatility. Summarizing

$$\ln(S_T/S_t) \sim N\left(r(T-t) - \frac{1}{2}\sigma^2(T-t), \sigma\sqrt{T-t}\right) \tag{2.60}$$

or

$$S_T/S_t \sim LN\left(r(T-t) - \frac{1}{2}\sigma^2(T-t), \sigma\sqrt{T-t}\right)$$

All these expressions enter d_1 and d_2 in the Black and Scholes formula. How can we calculate the probability that we exercise the call? The option is exercised if $S_T > K$. This reads for the continuous return $r_S = \ln(S_T/S_t)$

$$r_S = \ln(S_T/S_t) > \ln(K/S_t) \ .$$

But we know from (2.60) the distribution of r_S:

$$r_S = \ln(S_T/S_t) \sim N\left(r(T-t) - \frac{1}{2}\sigma^2(T-t), \sigma\sqrt{T-t}\right)$$

A calculation shows

$$P(S_T > K) = P\left(r_S > \ln(K/S_t)\right) = \Phi(d_2) \ .$$

The probability to exercise the option is equal to $\Phi(d_2)$.

Is there a derivation of the Black and Scholes equation which is more intuitive than the formal limit procedure in the CRR model? There are several ways to derive the pricing equation. One is using a pedestrian view of a trader - but to be honest this approach was only derived after more abstract derivations of the Black and Scholes equation existed. Another one, which is the original one of Black and Scholes, is to use replication in continuous time similar to the discrete model setup. The price to pay here is that one needs to know about stochastic calculus in continuous time.

2.5.2 Brownian Motion and Stochastic Calculus in a Nutshell

The theory of continuous time stochastic processes is a vast topic. Fortunately, for a user in finance only some simple rules need to be known to work with these mathematical objects. There two basic processes which are used to generate more complicated price processes: **Brownian motion** and the **Poisson process**. The first one is used to describe price evolutions where the price process follows a drift and a random walk (Brownian motion). But prices cannot jump - to consider this one uses the Poisson process and its generalizations. What is a Brownian motion?

Definition 2.5.1. *A standard Brownian motion W_t is a continuous time process where:*

1. $W_0 = 0$, *i.e. Brownian motion starts at zero.*
2. *For all $t > s > 0$ the increments $W_t - T_s$ are independent.*
3. $W_t - W_s$ *is normally distributed with mean zero and variance $t - s$, this leads to the classification 'standard'.*
4. *The paths of W_t are continuous, , i.e. for a fixed state ω the function $t \to W_t(\omega)$ is continuous.*

Intuitively, a Brownian motion price process can be seen as the results of many demand and supply activities. There are many of them which results in the normality assumption and there is no long term memory, the independence assumption. Although there are no jumps, Brownian motion is not a regular function of time - i.e. the paths of Brownian motion look jagged. There are several deep mathematical properties which follow from the definition of Brownian motion. First, although the paths are continuous, they are not differentiable. Hence, price changes have no velocity or a trend but are at each date completely random. Second, Brownian motion is a **Markov** process. The Markov property means that given the current value X_s, the prediction of future values of X_t is not improving if one considers the full history of the process compared to the information at time s only: The process only 'knows' its value at time s and has no memory about how it got there. Third, the paths of Brownian motion are of infinite first variation but of finite second or quadratic variation. The finiteness of the quadratic variation

$$[W, W](t) = t$$

reflects the basic assumptions in the construction of the Brownian motion: Squared differences in space of the process are proportional to linear differences in time. Intuitively, $[W, W](t)$ is the sum of the quadratic differences $(W(s) - W(v))^2$ over a time interval with length t, where interval is partitioned in finer and finer subintervals. See the Appendix 7.5 for more definitions of the first, quadratic variations and the proofs of the stated properties. Fourth, from the normality assumption follows that displacement in space is proportional to the root of time displacement, i.e.

$$(\Delta W_t)^2 = \Delta t \ , \ (dW_t)^2 = dt \tag{2.61}$$

where the second notation is a suggestive differential notation, see below. This different behavior of scaling in space and in time has far reaching consequences for the integral calculus and differential calculus of functions of Brownian motion. To see this, let $f(W)$ be a smooth function of Brownian motion. Taylor expanding the function gives

$$f(W) = \frac{\partial}{\partial W} f(W) dW + \frac{\partial^2}{2\partial W^2} f(W)(dW)^2 + \ldots .$$

The second order term is due to the scaling property:

$$f(W) = \frac{\partial}{\partial W} f(W) dW + \frac{\partial^2}{2\partial W^2} f(W) dt + \ldots .$$

That is, the chain rules has an additional term compared to classical analysis. If one considers the question to construct an integration theory of Brownian motion one faces the problem that Brownian motion paths are of unbounded first variation - i.e. we cannot us the 'usual' integral construction. But the second variation is finite. Using this one constructs an integral, the **Itô integral**. Without going in any details about the definition of differential and integral calculus we state the main properties which we need as workhorses. For ϕ, ψ two processes (strategies) which are 'nice enough'[12] the stochastic integral w.r.t. a **standard Brownian motion**

$$\int_0^t \phi(s,\omega)dW_s \tag{2.62}$$

is well defined. The integral should be thought of as a limit of the discrete sum

$$\sum_j \phi(t_j)(W(t_{j+1} - W(t_j)) \ .$$

The integral has the following properties. First, the integral is linear in ϕ (evident as a sum). Second, expectation of the integral is zero, i.e. $E\left[\int_0^t \phi(s,\omega)dW_s\right] = 0$. Since ϕ is adapted to the Brownian motion filtration the expected value of the sum is equal to the expected value of the two Brownian motions. This are zero by definition. Third, the **Itô isometry** holds:

$$E\left[\left(\int_0^t \phi(s)dW_s\right)^2\right] = \int_0^t E[\phi^2(s)]ds \ .$$

The Itô isometry follows from the independence of increments. That is, the square of the finite sum is decomposed into two parts: A part where the differences in Brownian motions are the same, i.e. where $(W(t_{j+1} - W(t_j))^2$ holds and the other terms of the form $(W(t_{j+1} - W(t_j))(W(t_{k+1} - W(t_k))$ are collected. The first ones scale as Δt and the second ones vanish due to the independence of increments. Fourth, the stochastic integral is an $\mathcal{F}_t$-martingale. This follows at once from the discrete version. Fifth, he quadratic covariation of a stochastic integral is given by:

$$\int_0^t \phi(s)\psi(s)[dW,dW]_s = \int_0^t \phi(s)\psi(s)ds \ .$$

Consider the martingale

$$Y_t = \int_0^t \phi(s)dW_s \ .$$

We use the following symbolic differential notation

$$dY_t = \phi_t dW_t \ .$$

[12] At each information time t the processes are known, i.e. they are adapted to the filtration and they are square integrable.

The expression dY_t is called a **stochastic differential** - its definition is given by the stochastic integral. This formal notation is useful. For the quadratic variation of a stochastic integral we get:

$$[dY, dY]_t = [\phi dW, \phi dW]_t = \phi_t^2 [dW, dW]_t = \phi_t^2 dt \tag{2.63}$$

where we used the Itô isometry and $[W, W](t) = t$. Since the quadratic variation of the Brownian motion equals t, we also have $d[W, W]_t = dt$. The rules are collected in the **multiplication table** :

$$(dW_t)^2 = dt \ , \ dtdt = 0 \ , \ dtdW = 0. \tag{2.64}$$

We apply these rules. Consider the stochastic differential equation (SDE)

$$dX_t = \mu(X, t)dt + \sigma(X, t)dW_t \ , \ X_0 \text{ given} \ , \tag{2.65}$$

which is the shorthand for

$$X_t = X_0 + \int_0^t \mu(X, s)ds + \int_0^t \sigma(X, s)dW_s \quad . \tag{2.66}$$

The equation (2.65) is a **diffusion**. This equation has two components. A random part generated by the Brownian motion and a drift part proportional to time.

Given a diffusion dX_t. What can we say about a transformed process $df(X, t)$ with f a smooth function? We perform a Taylor approximation, use the dynamics of X and the multiplication rule (2.64) to obtain:

$$\begin{aligned} dY &= \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial X}dX + \frac{1}{2}\left(\frac{\partial^2 f}{\partial t^2}(dt)^2 + \frac{\partial^2 f}{\partial X^2}(dX)^2 + 2\frac{\partial^2 f}{\partial X \partial t}dtdX\right) + o(dt^2) \\ &= \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial X}dX + \frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial X^2}dt + o(dt^2) \end{aligned}$$

This formal calculation shows that the term $\frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial X^2}(dX)^2$ is a **first order term**, i.e. he changes formally the ordinary chain rule. We summarize using the notation $\partial^n_{X_1,\ldots,X_n} f$ for the nth derivative of f w.r.t. to $X_1, \ldots, X_n$:

Proposition 2.5.2 (Itô's formula)**.** *Let f be a twice-continuous differentiable function in two variables and consider the SDE $dX_t = \mu_t dt + \sigma_t dW_t$. Then $f(X, t)$ has the differential*

$$df_t = \sigma \partial_X f dW_t + (\partial_t f + \mu \partial_X f + \frac{1}{2}\sigma^2 \partial^2_{XX} f)dt \tag{2.67}$$

or in integral form

$$f_t = f_0 + \int_0^t \partial_X f_X \sigma_s dW_s + \int_0^t \left(\partial_s f_s + \mu \partial_X f + \frac{1}{2}\sigma^2 \partial^2_{XX} f\right) ds \ . \tag{2.68}$$

The properties of the stochastic integral implies that f is a martingale if and only if the last integral in (2.68) vanishes, i.e. if f satisfies the following partial differential equation:

$$\partial_s f_s + \mu \partial_X f + \frac{1}{2}\sigma^2 \partial^2_{XX} f = 0 \ .$$

This gives a first indication that there are some deep mathematical relationships between stochastic calculus, analysis and the notion of martingales. The function $Z(x,t) = e^{ax-\frac{1}{2}a^2 t}$ is basic in the change of measure technique. Itô's formula implies that Z solves the SDE

$$dZ(W,t) = aZ(W,t)dW \ , \ Z(W,0) = 1 \ .$$

Hence Z is a martingale since W is a martingale and a stochastic integral w.r.t. Brownian motion is also a martingale.

Three main models in financial modelling are the Bachelier model, the geometric Brownian motion model (GBM) and the Ornstein-Uhlenbeck (OU) process. The definition and the main properties are given in Table 2.5 and 2.6. The solutions, the expected values and the variances are derived in Appendix 7.6.

Name	Process	Solution
Bachelier	$dX_t = \mu dt + \sigma dW_t \ , \ X_0 = x$	$X_t = x + \mu t + \sigma W_t$
GBM	$dX_t = \mu X_t dt + \sigma X_t dW_t \ , \ X_0 = x$	$X_t = xe^{(r-\frac{1}{2}\sigma^2)t + \sigma W_t}$
OU	$dX_t = -cX_t dt + \sigma dW_t \ , \ X_0 = x$	$X_t = e^{-ct}x + e^{-ct}\int_0^t e^{cs}\sigma dW_s$

Table 2.5: Bachelier model, the geometric Brownian motion model and the Ornstein-Uhlenbeck process

The Bachelier model leads to negative values for the process X. The geometric Brownian motion leads to positive prices X. This model is assumed in the Black and Scholes model. The Ornstein-Uhlenbeck model is considered for mean-reverting processes. The solution of the Bachelier model is trivial, the geometric Brownian motion model and the Ornstein-Uhlenbeck model both need a trick to find the solution - this is discussed in the Appendix 7.6.

Name	Expected Value	Variance
Bachelier	$E[X_t] = x + \mu t$	$\text{var}(X_t) = \sigma^2 t$
GBM	$E[X_t] = X_0 e^{\mu t}$	$\text{var}(X_t) = X_0^2 e^{2\mu t}\left(e^{\sigma^2 t} - 1\right)$
OU	$E[X_t] = e^{-ct}x$	$\text{var}(X_t) = \sigma^2 e^{-2ct}(e^{2ct} - 1)\frac{1}{2c} = \frac{\sigma}{2c}(1 - e^{-2ct})$

Table 2.6: Bachelier model, the geometric Brownian motion model and the Ornstein-Uhlenbeck process.

The expected value of the Ornstein-Uhlenbeck process goes to zero for t to infinity and c positive and the long term variance is constant, i.e. $var(X_t) \to \frac{\sigma}{2c}$ if time goes to infinity.

Consider two stochastic processes

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s \ , \ Y_t = Y_0 + \int_0^t \mu'_s ds + \int_0^t \sigma'_s dW_s$$

where $\mu, \mu', \sigma, \sigma'$ are stochastic processes satisfying appropriate conditions. The following **stochastic partial integration formula** holds:

Proposition 2.5.3.

$$X_t Y_t = X_0 y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + \int_0^t \sigma_s \sigma'_s ds \ . \tag{2.69}$$

The proof is given in Appendix 7.5. In differential notation the formula reads:

$$d(XY) = XdY + YdX + \sigma\sigma' dt \ . \tag{2.70}$$

As a corollary we state some useful formulae: We define the **covariation of two processes** $[X, Y]$ as

$$[X, Y]_t = X_t Y_t - X_0 Y_0 - \int_0^t Y_s dX_s - \int_0^t X_s dY_s \ . \tag{2.71}$$

This definition of the covariation is consistent with the quadratic variation $[W, W]_t = t$. From the multiplication table and integration by parts formula we deduce

$$d[X, Y]_t = [dX, dY]_t = \sigma\sigma' dt \ .$$

We next consider the **change of measure topic**. We write B_t for a risk free asset and X_t for a risky asset with dynamics

$$dB_t = rB_t dt, B_0 = 1 \ , \ dX_t = X_t(\mu dt + \sigma dW_t) \ , X_0 = 0 \tag{2.72}$$

where μ, σ and r are processes such that integrals w.r.t. Brownian motions are well-defined. The dynamics of the risky asset X is under a probability P. These two assets define the financial market.
Let $V^\xi = (B, X)$ be the portfolio process and $\xi = (\phi, \psi)$ the portfolio, i.e.

$$V_t^\xi = \phi_t B_t + \psi_t X_t$$

where we mark explicitly the portfolio. Such a portfolio is **self-financing** if a change in portfolio value is given by a change in asset prices only, i.e.

$$dV_t^\xi = \phi_t dB_t + \psi_t dX_t \ . \tag{2.73}$$

We write **discounted values** using the tilde, i.e. $\tilde{V} = V/N$ with N a numeraire. A numeraire is a strictly positive stochastic process. Let a portfolio ξ be a self-financing portfolio and $V^\xi = (B, X)$ the portfolio value process. What can be said about the discounted portfolio process $\tilde{V}^\xi = (\tilde{B}, \tilde{X}) = V^\xi/N$ under the portfolio ξ? Using Itô calculus we get:

Proposition 2.5.4. *ξ is a self-financing strategy for a portfolio V^ξ if and only if it is self-financing for the discounted portfolio process. If ξ is a self-financing strategy and X is a Q-martingale, then V^ξ is also a Q-martingale.*

A strategy is self-financing for V if and only if the same strategy is self-financing for $\tilde{V}$. The second statement seems innocent but it drives most of the theory which follows. The notion of arbitrage is unchanged compared to the discrete time model. A key and simple to prove property is:

Proposition 2.5.5. *Suppose that there exists a self-financing strategy with portfolio dynamics*

$$dV = kV dt \ .$$

If the market is arbitrage free, k equals the risk free rate r.

With more than one risk less asset arbitrage is possible. In discrete time models the existence of a risk neutral probability or an equivalent martingale measure Q is equivalent to the absence of arbitrage. This hold true also in continuous time:

Proposition 2.5.6 (First Fundamental Theorem of Finance)**.** *A market is arbitrage free if and only if there exists an equivalent martingale measure Q.*

We recall that two equivalent probability measures are related by the density process Z_t:

$$Q_t(A) = \int_A Z_t(\omega) dP_t(\omega) \ , \ A \in \mathcal{F}_t \ . \tag{2.74}$$

The volume A measured with a stick Q is equal to using a different stick P if corrections by the stretching factor Z are used. The process $Z = (Z_t)_t$ is the **density process** or Radon-Nikodym derivative and we write in differential notation

$$Z_t(\omega) = \frac{dQ_t(\omega)}{dP_t(\omega)} \ . \tag{2.75}$$

We summarize:

The change of measure leads to a re-weighting of probabilities of the price paths of the risky assets. Neither are paths created or deleted. This change of measure technique affects only the drift of diffusion processes. The volatility structure is not affected by such a measure change. Therefore, the technique is the tool to transform processes which are not martingales into martingales. To change the volatility structure of a diffusion one needs to use different methods such as the time change method. Using this method one alters the number of paths, i.e. the change of measure are not of an absolutely continuous type. The notions of self-financing and no arbitrage are invariant under the choice of the numeraire. The choice of a numeraire is a matter of convenience and is dictated by the valuation problem at hand.

We consider the following questions:

- What happens to a portfolio if we change the numeraire?
- How does a Brownian motion changes if we change to a new equivalent measure? The answer is the Cameron-Martin-Girsanov Theorem.

What happens if we change the numeraire from N to N'? The Cameron-Martin-Girsanov (CMG) Theorem, see below, states that the exists a new martingale measure Q' such that $\frac{V(T)}{N('T)}$ is a Q'-martingale, i.e.

$$\frac{V(s)}{N'(s)} = E^{Q'}[\frac{V(T)}{N('T)}|\mathcal{F}_s] \ . \tag{2.76}$$

Using this representation and Proposition 2.4.6 implies:

$$\begin{aligned}\frac{V(s)}{N'(s)} &= E^{Q'}[\frac{V(T)}{N'(T)}|\mathcal{F}_s] \\ &= Z_s^{-1}E^{Q}[Z_T\frac{V(T)}{N'(T)}|\mathcal{F}_s] \\ &= \frac{dQ}{dQ'}(s)E^{Q}[\frac{dQ'}{dQ}(T)\frac{V(T)}{N'(T)}|\mathcal{F}_s] \\ &= \frac{N(s)}{N'(s)}E^{Q}[\frac{N'(T)}{N(T)}\frac{V(T)}{N'(T)}|\mathcal{F}_s] \ .\end{aligned}$$

Summarizing:

$$\boxed{V(s) = E^{Q'}[\frac{V(T)}{N'(T)}|\mathcal{F}_s] = E^{Q}[Z_T\frac{V(T)}{N(T)}|\mathcal{F}_s] \ , \ Z_T = \frac{N(s)}{N'(s)}\frac{N'(T)}{N(T)} \ .} \tag{2.77}$$

Examples

1. **Spot and forward measure**. The spot numeraire reads $N(t) = B(t) = e^{\int_0^t r(s)ds}$ and the forward measure numeraire is $N(t,T) = p(t,T)$, i.e. the zero coupon bond for maturity T. We elaborate on the second example for a fixed date T.

   $$\frac{dQ^T}{dQ} = \frac{1}{B(T)p(0,T)}$$

   defines an equivalent probability Q^T to the probability Q associated to the numeraire $B(t)$. For $t \leq T$, we have

   $$\frac{dQ^T}{dQ}(t) = E^Q\left[\frac{dQ^T}{dQ}(T)|\mathcal{F}_t\right] = \frac{p(t,T)}{B(t)p(0,T)} \ .$$

   Q^T is called the forward measure. Consider the zero bond price ratio $p(t,S)/p(t,T)$ with $S > T$. This ratio is a Q^T-martingale. To prove this we use Bayes' rule for

$w < t \leq \min(S,T)$:

$$E^Q\left[\frac{p(t,S)}{p(t,T)}|\mathcal{F}_w\right] = \frac{E^Q\left[\frac{p(t,T)}{p(0,T)B(t)}\frac{p(t,S)}{p(t,T)}|\mathcal{F}_w\right]}{\frac{p(w,T)}{B(w)p(0,T)}}$$
$$= \frac{\frac{p(w,S)}{B(w)}}{\frac{p(w,T)}{B(w)}} = \frac{p(w,S)}{p(w,T)} . \tag{2.78}$$

This result and

$$1 + \alpha F(t,S,T) = \frac{p(t,S)}{p(t,T)} .$$

imply that the forward LIBOR rate $F(t,S,T)$ is a Q^T-martingale.

2. Let X_T be an option payoff with fair price $C(t,X)$ in a model setup with interest rate risk:

$$C(t,X) = E^Q\left[e^{-\int_t^T r(s)ds}X_T|\mathcal{F}_t\right] .$$

To calculate the price one has to solve for a double integral, i.e. an integral for the r and X risk factors. Only if r and X are independent random variables, the integration splits into two one-dimensional integrals:

$$C(t,X) = E^Q\left[e^{-\int_t^T r(s)ds}|\mathcal{F}_t\right]E^Q\left[X_T|\mathcal{F}_t\right] = p(t,T)E^Q\left[X_T|\mathcal{F}_t\right] .$$

Unfortunately, the two random variables are dependent and a double integral has to be considered under the measure Q. But the change or measure technique allow us to reduce the number of integrations as follows. The idea is to use the T-zero bond as a numeraire and define the T-forward measure Q^T by the requirement that the process $C_t/p(t,T)$ is a Q^T-martingale. Since the zero bond pays one at maturity, $p(T,T) = 1$, and $C(T,X) = X_T$ we get with the martingale property

$$\frac{C(t,X)}{p(t,T)} = E^{Q^T}\left[X_T|\mathcal{F}_t\right] ,$$

or

$$C(t,X) = p(t,T)E^{Q^T}\left[X_T|\mathcal{F}_t\right] .$$

This is the desired reduction of a double integral to a single integration.

We consider the **Cameron-Martin-Girsanov (CMG) Theorem**.

Proposition 2.5.7 (Cameron-Martin-Girsanov (CMG))**.** *Let W_t be a P standard Brownian motion , $\nu_t \in \mathcal{F}_t$[13] and $E_P[e^{\frac{1}{2}\int_0^T \nu^2(s)ds}] < \infty$. There exists a measure $Q \sim P$ with density*

$$Z_T = \frac{dQ}{dP}(T) = e^{-\int_0^T \nu(s)dW_s - \frac{1}{2}\int_0^T \nu^2(s)ds} \tag{2.79}$$

and the process $\tilde{W}_t = W_t + \int_0^t \nu(s)ds$ is a Q- standard Brownian motion.

[13] The filtration is generated by the P-Brownian motion.

Equivalently, $\tilde{W}$ is a Brownian motion under P with drift $-\nu_t$. If we correct the original Brownian motion W by the integrated process we get a standard Brownian motion $\tilde{W}$ under the equivalent measure Q. Why is this useful? Consider an option pricing model where the discounted price process under P is not a martingale due to the presence of a drift term (the expected value of the process is not constant). To apply the option pricing formula we need to change the original process into a new one which is a martingale. That is we have to remove the drift or equivalently replace the W-term by the $\tilde{W}$-term.

The integrability condition is called the Novikov condition. It implies that the solution

$$Z_t = Z_0 e^{\int_0^t \nu(s)dW(s) - \frac{1}{2}\int_0^t \nu^2(s)ds}$$

of the equation $dZ_t = \nu_t Z_t dW_t$ is a martingale. Applying Itô's formula to the density we get:

$$dZ_t = Z_t(\nu_t dW_t - \frac{1}{2}\nu_t^2 dt + \frac{1}{2}\nu_t^2 dt) = Z_t \nu_t dW_t \ , \ Z_0 = 1 \ . \tag{2.80}$$

Since Z is proportional to a Brownian motion and ν_t satisfies the integrability condition it follows that Z **is a P-martingale.**

To gain some intuition for the density Z's functional form, we consider two examples. Let X be a normal distributed random variable with mean μ and unit variance. We write $E_\mu[f(X)]$ with f an arbitrary function for the expected value under such a μ-mean distribution. We want to express this expected value using a distribution with mean zero, i.e. using $E_0[f(X)]$:

$$\begin{aligned} E_0[f(X)] &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-\frac{1}{2}x^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-\frac{1}{2}(x-\mu)^2} e^{\mu^2 - \mu x} dx \\ &= E_\mu[f(X) e^{\mu^2 - \mu X}] \ . \end{aligned}$$

This shows how the exponential term appears in the density Z in the CMG Theorem when a drift correction is made. As a second example, consider CMG for a Brownian motion. For X a Gaussian with mean μ and variance σ^2 the Gaussian integral $E_P[e^{aX}]$ equals[14]

$$E_P[e^{aX}] = e^{a\mu + \frac{1}{2}\sigma^2 a^2} \ .$$

If $X = W$ is a standard Brownian motion under P, i.e. W_T is normal with mean zero and variance T under P, we have

$$E_P[e^{aW_T}] = e^{\frac{1}{2}\sigma^2 a^2} \ .$$

We define the Radon-Nikodym density at the final time T by

$$\frac{dQ}{dP}|_T = e^{-\nu B_T - \frac{1}{2}\nu^2 T} \ , \tag{2.81}$$

[14] To derive this result one needs to perform Gaussian integrals.

for an equivalent measure Q. Then,

$$E_Q[e^{aW_T}] = E_P[\frac{dQ}{dP}|_T e^{aW_T}] = E_P[e^{-\nu W_T - \frac{1}{2}\nu^2 T} e^{aW_T}] = e^{-a\nu T + \frac{1}{2}a^2 T} \ .$$

Hence, $E_Q[e^{aW_T}]$ is a Gaussian with mean $-\nu T$ and variance T. This holds for all $t < T$: W_t is a standard Brownian motion under P and a Brownian motion with constant drift ν under Q. The process $\tilde{W}_t = W_t + \nu t$ is again a Q standard Brownian motion.

We apply the theorem to the price dynamics of the risky asset used in **Black and Scholes.** Let $dS/S = \mu dt + \sigma dW$ be a geometric Brownian motion under P Using Itô's formula for the discounted process $\tilde{S} = S/B$ we get

$$d\tilde{S} = \tilde{S}(\mu - r)dt + \tilde{S}\sigma dW \ .$$

This process is a martingale if and only if $\mu = r$. To find the martingale measure Q we use $\tilde{W}_t = W_t + \int_0^t \nu(s)ds$ which leads to (ν is constant)

$$d\tilde{S} = \tilde{S}(\mu - r + \sigma\nu)dt + \tilde{S}\sigma d\tilde{W} \ .$$

$\tilde{S}$ **is a Q-martingale** if the drift term vanishes, i.e.

$$-\nu = \frac{\mu - r}{\sigma}$$

holds. ν is the negative value of the **market price of risk.** We could equally consider the underlying asset S as our numeraire. Then B/S has to be a martingale. Applying Itô we have:

$$d\left(\frac{B}{S}\right) = \frac{(r + \sigma^2 - \mu)B}{S}dt - \sigma\frac{B}{S}dW \ .$$

To obtain a martingale, the drift has to vanish, i.e. $\mu = r + \sigma^2$. With the Girsanov theorem the dynamics of the two assets in the martingale measure related to the S-numeraire reads

$$dS = (r + \sigma^2)Sdt + \sigma S dW \ , \ dB = Brdt \ .$$

The new process $\tilde{W}$ which makes B/S a martingale requires a correction $\nu = \frac{r+\sigma^2-\mu}{\sigma}$. Pricing a call option under this measure has to give the same value than using the traditional risk less asset as numeraire (good exercise).

2.5.3 Derivation of Black and Scholes using Stochastic Calculus

The model of Black and Scholes starts with a dynamics of a single risky asset S and a risk less one B. This two assets define the financial market. There are no transaction costs, the instruments can be traded freely and instantaneously either long or short at the price quoted. We assume that S is driven by a single source of risk (which we model below as a Brownian motion) and we use the notions of arbitrage, complete markets and hedging in continuous time. Figure 2.17 illustrates the main results for continuous time

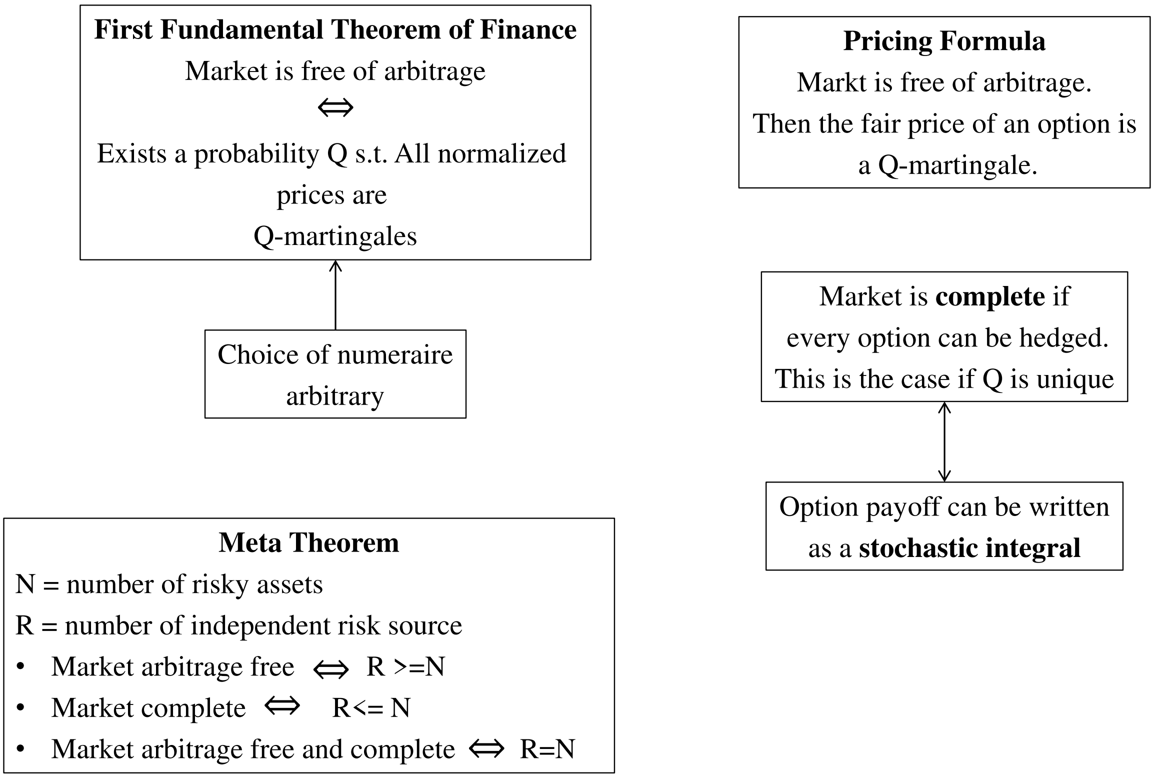


Figure 2.17: Main results of continuous time finance. *Source: T. Björk (2009)*

finance. Since their logic is not different from their discrete time counterparts we do not discuss them. A **portfolio** in the Black and Scholes model is $\xi_t = (\phi_t, \psi_t)$ where ϕ represents the amount of the risk less asset and ψ the number of shares of the risky asset. The **portfolio value** is

$$V_t^\xi = \phi_t B_t + \psi_t S_t \ .$$

We **always** consider self-financing strategies. By the Meta theorem, see Figure 2.17, the Black and Scholes market is arbitrage free and complete, i.e. there exists a single **risk neutral probability measure (martingale measure)** Q such that the discounted price process $\tilde{S} = S/B$ is a martingale. The risky asset S_t satisfies under the physical measure P

$$dS_t = S_t \mu dt + \sigma S_t dW_t \ , \ S_0 \text{ given} \tag{2.82}$$

where μ and σ are constants (the drift and the volatility, respectively) and W_t is a standard Brownian motion under the objective probability P. The solution of the stochastic differential equation (SDE) for the risky asset is (see Appendix 7.5, (7.15)):

$$S(t) = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t} \ . \tag{2.83}$$

The risk less asset price grows as:

$$dB_t = r B_t dt \ , \ B_0 = 1, \text{ i.e. } B_t = e^{rt} \ , \tag{2.84}$$

with r the risk less interest rate. S is **not** a martingale under P since $E^P[S(t)] = S_0 e^{\mu t}$ which is not constant. The discounted prices $\tilde{S}$ satisfies, see Appendix 7.5, (**??**):

$$E^P[\tilde{S}(t)] = S_0 e^{\mu t - rt} \ . \tag{2.85}$$

To satisfy the martingale property we need $\mu = r$. That for, we We remove the drift from the W_t Brownian motion under P using the CMG Theorem: There exists a new Brownian motion $\tilde{W}$

$$d\tilde{W}_t = \nu dt + dW_t \ , \ \nu = \frac{r - \mu}{\sigma} \tag{2.86}$$

such that $\tilde{S}$ with the Q dynamics

$$d\tilde{S} = \sigma \tilde{S} d\tilde{W}_t \ . \tag{2.87}$$

is a martingale under this new measure (see Appendix 7.5). Using the representation of self-financing strategies

$$\tilde{V}_t^\xi = V_0 + \int_0^t \psi_s d\tilde{S}_s$$

and the martingale representation theorem it follows that the discounted value process is a Q-martingale, i.e. ,

$$\tilde{V}_t^\xi = E_Q[\tilde{V}_T^\xi | \mathcal{F}_t] \ .$$

This is equivalent to

$$V_t^\xi = E_Q[e^{-r(T-t)}\tilde{V}_T^\xi|\mathcal{F}_t] \ . \tag{2.88}$$

Now we are done to derive the Black and Scholes formula - we have to calculate the expected value in the last formula which gives the call option pricing formula (2.55), see Appendix 7.5 for the calculation.

2.6 Hedging, Risk Management and P&L

We **always work** in this section either in the discrete binomial CRR or the Black and Scholes model.

2.6.1 Introduction to the Greeks

Greeks are sensitivities of option prices with respect to parameters (partial derivatives). **Delta** (Δ) is a number that measures how much the theoretical value of an option will change if the underlying stock moves up or down by 1 in a given currency. A Delta of +0.5 implies that if an underlying stock rises by CHF 1, the theoretical option price increases by CHF 0.5. The sum of the absolute values of the Deltas for an European call and an European put with the same strike and maturity is 1. This follows from put-call parity by taking the partial derivative w.r.t. to S. Delta is a linear operation: the Delta of a portfolio equals the sum of the constituent's Delta. This allows to define the **position Delta**, i.e.

$$\text{Position Delta} = \Delta \times \text{Quantity} \times \text{Number Shares} \quad ,$$

where Δ is the option theoretic Delta, Quantity' is the quantity of option contracts and 'Number Shares' is the number of shares of stock per option contract.

An investor is long 10 option contracts of 50-calls (i.e. strike 50) on Nestle stock with a Delta of 0.5 with 70 shares of the stock per option contract and the investor is short 200 Nestle stocks. The position Delta is from the stock and option part:

$$-200 + 0.5 \times 10 \times 70 = +150 \ .$$

Theoretically a change of Nestle stock by CHF 1 leads to a gain/loss of CHF 150 in this portfolio.

The Delta for ATM is close to 0.5 but not exactly (why?). Delta is not uniformly strong for ATM, ITM and OTM options - Gamma measures the Delta sensitivity.

Gamma Γ is an estimate of how much the Delta of an option changes when the price of the stock moves. A big Gamma means that the Delta can start changing strongly for even a small move in the stock price. A positive Gamma means that the Delta of long calls will become more positive and move toward +1 when the stock prices rises, and less positive and move toward zero when the stock price falls. The Gamma as a second derivative is also a linear operation and we therefore have for the position Gamma:

$$\text{Position Gamma} = \Gamma \times \text{Quantity} \times \text{Number Shares} \quad ,$$

where Γ is the option theoretic Gamma. The graph of a call option shows that the slope (the Delta) is always positive and monotone increasing from zero to 1. The increase is highest close to ATM, i.e. the Gamma has a peak value in the ATM-region and goes monotonically to zero if either deep ITM or OTM regions are considered where the Delta becomes almost stable. Gamma is the same for call and puts. The Gamma of ATM options is higher when either volatility is lower or there are fewer days to expiration. But if an option is sufficiently OTM or ITM, the Gamma is also lower when volatility is lower or there are fewer days to expiration. What this all means to the option trader is that a position with positive Gamma is relatively safe, that is, it will generate the Deltas that benefit from an up or down move in the stock. A position with negative Gamma can be dangerous. It will generate Deltas that will hurt the trader in an up or down move in the stock. But all positions that have negative Gamma are not all dangerous. For example, a short straddle and a long ATM butterfly both have negative Gamma. But the short straddle presents unlimited risk if the stock price moves up or down. The long ATM butterfly will lose money if the stock price moves up or down, but the losses are limited to the total cost of the butterfly.

Although the sensitivities are given by partial derivatives of the option pricing formula it is meaningful to consider Delta and Gamma hedging in a simple, discrete setup. We consider a call option with strike 95$ and underlying with price 100$ today. The option has a market price of 8.5$. But as the option's theoretical-model price is 8.75$, an investor may be tempted to buy the call and make a gain of 0.25$. This is, however, a risky strategy. We show how the mispricing effect can be cashed-in without any risk. To simplify the analysis, assume that the underlying value can only move in steps of ±$5. Table 2.7 gives the price of the option today and tomorrow under both the up- and down-scenarios.

Today			Tomorrow, down			Tomorrow, up		
S	Prob.	Price	S	Prob.	Price	S	Prob.	Price
70	0.08	-	65	0.08	0	75	0.08	-
85	0.17	-	80	0.17	0	90	0.17	-
100	0.50	2.50	95	0.50	0	105	0.50	5.00
115	0.17	3.33	110	0.17	2.5	120	0.17	4.17
130	0.08	2.92	125	0.08	2.5	135	0.08	3.33
Sum	1.00	8.75	Sum	1.00	5.00	Sum	1.00	12.50

Table 2.7: The theoretical option price under different scenarios (K =$95).

If the underlying's price decreases (increases), the option price becomes \$5 (\$12.5). With the risky strategy, i.e. by buying the option with the lower market price and hoping to gain \$0.25, we lose 3.5\$ if prices fall and win \$4 if prices rise. Using Δ, we can cash in the \$0.25 with a risk less strategy. In either scenario Δ is given by

$$\Delta = \frac{8.75 - 5}{100 - 95} = \frac{12.5 - 8.75}{105 - 100} = 0.75,$$

i.e. the Δ is the same. Given this result, we set up the following strategy:

- sell 0.75 units of the underlying asset today at the price of \$100;
- buy the call for its market price of \$8.5;
- in the following period, sell the call, no matter what the value of the underlying asset is.

We check that this is a risk less strategy us to cash in the mispricing amount:

- Scenario down: We lose \$3.5 on the call and gain $0.75 * (100 - 95) =$\$3.75 from buying the underlying asset. The net gain is equal to \$0.25.
- Scenario up: We gain \$4 on the call and lose $0.75*(100-105) =$\$-3.75 from buying the underlying asset. The net gain is equal to \$0.25.

Therefore, Δ is the hedge ratio required to keep hold of 0.25\$. In the second step, we relax the assumption of a **constant** Δ, i.e. we bring the Gamma into play. We have a call option with strike \$100 and price \$5 today. In the up- and down-scenarios the option price becomes \$2.5 and \$12.5, respectively. We consider underlying price movements by ± 10 in one period. The discrete time Δ under the two scenarios is given by

$$\Delta^- = \frac{5 - 2.5}{100 - 90} = 0.25, \quad \Delta^+ = \frac{12.5 - 5}{110 - 100} = 0.75.$$

Hence, Δ is not constant. Which Delta do we have to choose? The goal is to fix Δ in such a way that

$$\Delta(100) \cdot \delta S - \Gamma(100) \cdot (\delta S)^2 = 0,$$

i.e. the Γ-corrected Δ provides us with a position which does not depend on the underlying's risk. To achieve this, we compute Γ as the change in Δ over the change in the underlying price,

$$\Gamma = \frac{0.75 - 0.25}{110 - 90} = 0.025.$$

Therefore,

$$\Delta(100) \cdot 20 - 0.025 \cdot (20)^2 = \Delta(100) \cdot 20 - 10 = 0;$$

$$\Delta(100) = \frac{10}{20} = 0.5.$$

If we take Δ equal to 0.5, we make a risk less gain. Suppose the underlying's price falls to \$90. Then the option is worth \$2.5 and we make a loss of $5 - 2.5 =$\$2.5 by selling the option. Selling 0.5 units of the underlying asset gives us a gain of $0.5(100 - 90) =$$5 and the net gain is $2.5. Suppose now the underlying's price increases to $110. Then the option is worth $12.5 and the gain from selling it is $12.5 - 5 =$$7.5. Selling 0.5 units of the underlying asset gives us a loss of $0.5(100 - 110) =$$-5 and the net gain is $2.5. Summarizing, if the underlying's price falls, we are 25 Δ-units overhedged. If the underlying's price rises, we are underhedged by the same amount. Γ is on average equal to 2.5$ Δ-units for one point movement in the underlying's price. Correcting our Δ-hedge exactly by this amount provides us with a risk less strategy. Using this correction of the Delta one can then risk less cash-in mispricings.

Theta Θ, or time decay, is an estimate of how much the theoretical value of an option decreases when 1 day passes. The Thetas for a identical calls and puts are not equal. The difference in Theta between calls and puts depends on the cost-of-carry for the underlying stock. When the dividend yield is less than the interest rate - the cost-of-carry for the stock is positive - Theta for the call is higher than for the put. The difference between the extrinsic value of the option with more days to expiration and the option with fewer days to expiration is due to Theta. Therefore, long options have negative Theta and short options have positive Theta. If options are continuously losing their extrinsic value, a long (short) option position will lose (gain) money because of Theta. Theta value does not decreases linearly over time since the value is not linear distributed between OTM, ATM and ITM. If volatility increases the value of the option is more varying over time which reduces the Theta. Gamma and negative Theta are dual to each other: If Gamma is highest for a long call position which will make money if there is a large move in the underlying, negative Theta is also largest. Therefore, if the stock move is not realized, Theta will negatively affect the position value. Position Theta is calculated in the same way as position Delta, but instead of using the number of shares of stock per option contract, Theta uses the dollar value of 1 point for the option contract.

Consider a long call option on CHFUSD with a Delta of 30 percent for one million USD with a Gamma of 3 percent. To hedge this position on sells 300'000 in the spot. If the spot first moves up by 1 percent and then falls back by 1 percent in the first step one has to sell additional shares and then buy back the same amount of shares to remain Delta neutral in these two steps. Although the amount of shares is the same in both transactions, they were sold at a higher price than bought - this is the long Gamma position, which leads to a gain. For a long put and long share position a similar analysis holds. Consider to buy a 40 Delta EURUSD put[15] for one million with a Gamma of 5.8% you can hedge the option position by buying 400,000 in the spot market. If the spot increases by one percent the option becomes a 34.2 Delta Put ($-40 + 5.8 = -34.2$). Since you have a long position in the spot of 400,000 this implies that you become long in Delta and must sell 58,000 of your position in the spot to remain Delta neutral. Similarly, if the spot descends to its original position you will have to buy back the 58,000 that you

[15]The Delta of the Put is actually -40 since the Delta of a long position in a Put is negative. However, market convention is to omit the negative sign.

just sold.

We consider Delta and Gamma Hedging in for the portfolio V:

- Short $1'000$ calls, Time-to-Maturity (TtM) 90 days, strike 60, volatility 30%, risk less rate 8%. The currency is irrelevant.
- The fair option price using Black and Scholes is 4.14452 and the Delta is 0.581957. We therefore receive a premium of 4144.52 by selling the options.
- To hedge the position we buy 581.96 stocks for the price 60. That for we borrow (cash)
$$581.96 \times 60 - 4144.52 = 34'917.39 - 4144.52 = 30'772.88 \ .$$

The portfolio value today is zero. We consider the portfolio value after 1 day, i.e. TtM is 89 days.

In the *Scenario 'unchanged'* the underlying value remains at 60. Using Black and Scholes, the option is worth 4.11833 (Theta, i.e. time value). Therefore the option liability value is lower but the cash liability increased:

$$30'779.62 = 30772.88 * (1 + 0.08/365) \ .$$

Therefore, a gain 19.44 follows, see Table 2.8. The result for the two other scenarios 'up' and 'down' are calculated in the same way and reported in the table.

	Value		
	unchanged	**up**	**down**
Underlying	34'917.39	35'499.35	34'335.44
Cash	-30'779.62	-30'779.62	-30'779.62
Option	-4'118.33	-4'721.50	-3'559.08
Sum	19.44	-1.77	-3.26

Table 2.8: Value of the portfolio V after 1 day for different scenarios.

This shows that the Delta hedge is effective for small changes in the underlying value.

We continue the example requiring that we also want to hedge the Gamma. Since one option is used for the Delta hedge, we need a second option to achieve also Gamma neutrality. The data of this option are:

- Call, TtM 60 days, strike 65.
- All other parameters are the same as for the first option, see Table 2.10.

	TtM	Strike	Option Price	Delta	Gamma
Option 1	90/365	60	4.14452	0.581957	0.043688
Option 2	60/365	65	1.37825	0.312373	0.048502

Table 2.9: Option data.

We require the portfolio V to be both Delta and Gamma neutral. That for we choose the number of stocks and the number z of option 2 such that the conditions have to hold:

$$\Delta V = x - 1'000\Delta_{\text{Opt1}} + z\Delta_{\text{Opt2}} = 0$$

and

$$\Gamma V = -1'000\Gamma_{\text{Opt1}} + z\Gamma_{\text{Opt2}} = 0 \ .$$

Solving these two linear equations gives

$$x = 300.58 \ , \ z = 900.76$$

To fix cash, one solve $V = 0$ at time 0, i.e.

$$V = xS + \text{Cash} - 1000 * \text{Opt1} + z * \text{Opt2} = 0$$

gives

$$\text{Cash} = -15'131.77 \ .$$

To be Delta and Gamma neutral we are long in the underlying, long in option 2 and short cash. The following table compares the hedge effectiveness between Delta and Delta & Gamma hedging.

Underlying after 1d	**Delta & Gamma**	**Delta**
58	-2.04	-71.35
58.5	0.3	-31.56
59	1.07	-3.26
59.5	0.81	13.69
60	0.02	19.45
60.5	-0.79	14.22
61	-1.11	-1.77
61.5	-0.49	-28.24
62	1.52	-64.93

Table 2.10: Delta & Gamma vs. Delta Hedge.

Vega is an estimate of how much the theoretical value of an option changes when volatility changes by 1 percent. Higher volatility means higher changes to realize gains, the higher risks to make losses are not there since we consider options and not forwards. Therefore, higher option prices follow. Positive Vega means that the value of an option

Position	Delta	Gamma	Theta	Vega	Rho
Long stock	1	0	0	0	0
Short stock	-1	0	0	0	0
Long call	positive	positive	negative	negative	positive
Short call	negative	negative	positive	negative	negative
Long put	negative	positive	negative	negative	negative
Short put	positive	negative	positive	negative	positive

Table 2.11: Relation between positions and their Greeks.

position increases when volatility increases, and decreases when volatility decreases. For a 50-call with price CHF 1, a Vega of $+0.5$ and a volatility of 20 percent an increase of the volatility to 31 percent leads to a option price of CHF 1.5. Vega is highest for ATM options, and is progressively lower as options are ITM and OTM. Position Vega measures how much the value of a position changes when volatility changes 1.00%. Position Vega is calculated in the same way as position Theta.

Rho ρ is an estimate of how much the theoretical value of an option changes when interest rates move 1.00 percent. The Rho for a call and put at the same strike price and the same expiration month are not equal. Rho is one of the least used Greeks. When interest rates in an economy are relatively stable, the chance that the value of an option position will change dramatically because of a drop or rise in interest rates is pretty low. Nevertheless, we'll describe it here for your edification. Long calls and short puts have positive Rho. Short calls and long puts have negative Rho. How does this happen? The cost to hold a stock position is built into the value of an option. Since an option can be replicated by a stock and loan position, stocks and options are substitutes of each other. Due to the leverage effect to replicate the value of a stock position much less has to be spent for the corresponding options. Put it different to build up an option position one has to borrow money to buy the shares. This interest costs are part of the option price and the more expensive it is to hold a stock position, the more expensive the call option.

Table 2.11 summarizes relations between positions and their Greeks.

In the Black and Scholes model the sensitivities can be calculated explicitly. They are shown in next table.

Sensitivity w.r.t.	Math	Finance	Expression
Underlying S	$\frac{\partial C(S)}{\partial S}$	Delta Δ	$\Delta_C = \Phi(d_1) > 0$
	$\frac{\partial P(S)}{\partial S}$		$\Delta_P = \Phi(d_1) - 1 < 0$
Time-to-maturity τ	$\frac{\partial C(\tau)}{\partial \tau}$	Theta Θ	$\Theta_C = -S\sigma\phi(d_1)/(2\sqrt{\tau}) - rKe^{-r\tau}\Phi(d_2) < 0$
	$\frac{\partial P(\tau)}{\partial \tau}$		$\Theta_P = -S\sigma\phi(d_1)/(2\sqrt{\tau}) + rKe^{-r\tau}\Phi(-d_2) < 0$
Risk free rate r	$\frac{\partial C(r)}{\partial r}$	Rho ρ	$\rho_C = Ke^{-r\tau}\tau\Phi(d_2) > 0$
	$\frac{\partial P(r)}{\partial r}$		$\rho_P = -Ke^{-r\tau}\tau\Phi(-d_2) < 0$
Vola σ	$\frac{\partial C(\sigma)}{\partial \sigma}$	Vega ω	$\omega_C = \phi(d_1)S\sqrt{\tau} > 0$
	$\frac{\partial P(\sigma)}{\partial \sigma}$		$\omega_P = \omega_C$
Underlying S	$\frac{\partial^2 C(S)}{\partial S^2}$	Gamma Γ	$\Gamma_C = \phi(d_1)/(S\sigma\sqrt{\tau}) > 0$
	$\frac{\partial^2 P(S)}{\partial S^2}$		$\Gamma_P = \Gamma_C$

2.6.2 Relating the Greeks via Black and Scholes

We discussed the sensitives as if they were independent. But they are not. That for we discuss the derivation of the pricing equation of Black and Scholes from a pedestrian view. This discussion is taken from Neftci (2008). Consider a trader buying an at-the-money call option with price C on some underlying stock with price S. The trader may fund his position by either shorting the underlying or by taking a loan from the money market desk at a fixed, non-zero funding cost r. Figure 2.18 shows that the net position of the market maker is risky.

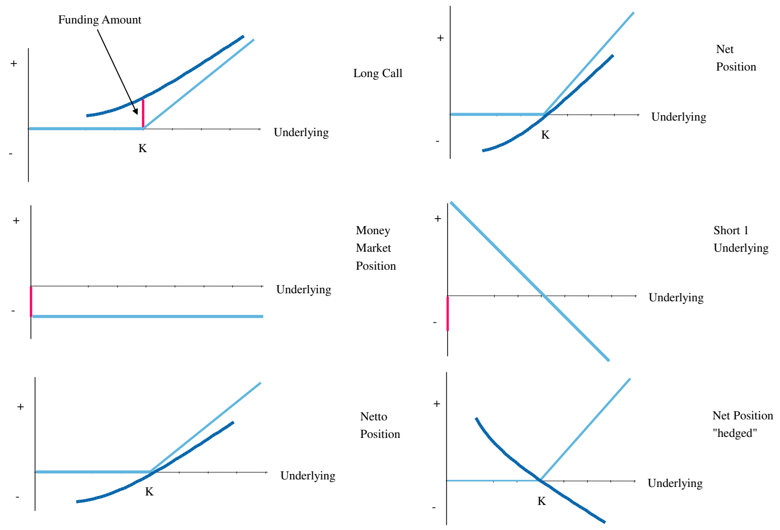


Figure 2.18: Left Panel: Net position market maker. Right Panel: Net position where the market maker is short 1 underlying.

The market maker makes a profit if the underlying price increases and loses in the other case. But the trader also needs to manage his risk. One idea is to short one unit of the underlying since a short position in the underlying value gains if the underlying falls. Hence a **finite change** from S by δS leads to change in the short position by $-\delta S$. At this point the non-linearity of the call payoff and the linearity of the underlying payoff matter: Any change of the underlying value has the same slope sensitivity of $+1$, but the call sensitivity measured with the slope is bounded by 0 and $+1$. A small change in the option price is proportional to the underlying value, i.e.

$$\delta C = C(S + \delta S) - C(S) = \Delta \delta S + \text{ Error } .$$

By going short only Δ of the underlying we get for the sensitivity of the net position.

$$\delta V = \Delta \delta S - \Delta \delta S + \text{ error } = \text{ error } .$$

But the Delta hedge over time has to be adjusted. This is a second order effect. Figure 2.19 shows what happens in a simple scenario. We assume that the underlying value oscillates with an annual volatility σ around the mid point S^0. The underlying can reach two states in the oscillation S^- or S^+, where in the case of a geometric Brownian motion we have:

$$S^+ = S^0 + \delta S = S^0(1 + \sigma\sqrt{(T - t)}) .$$

Hence δS is the percentage oscillation of the underlying which is proportional to the square root of time t.

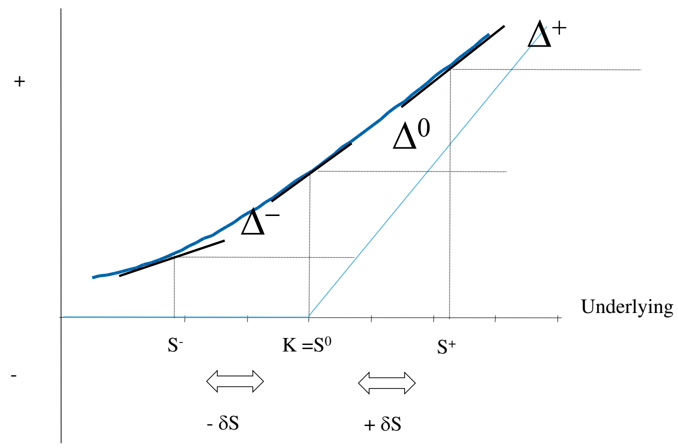


Figure 2.19: Delta hedge over time.

Consider the loop S^0 to S^+ and back. The Delta Δ^+ is larger than the original one at 0, i.e. the market maker has to increase his short position. If the underlying value drops back to the initial price, the position has to reduced by the market maker. For such a two-period oscillation $S^0 \to S^+ \to S^0$ the Delta hedge generates the cash gain:

$$\Delta^+(S^+ - S^0) - \Delta^0(S^+ - S^0) = (\Delta^+ - \Delta^0)\delta S > 0 \ .$$

For the cycle $S^0 \to S^- \to S^0$ also a gain follows. These cycles are specific. How does the result changes if only small but arbitrary moves are allowed? To see this, we transform the gain into a gain rate, i.e.

$$(\Delta^+ - \Delta^0)\delta S = \frac{(\Delta^+ - \Delta^0)}{\delta S}(\delta S)^2 \ .$$

This is the two-period gain. The gain in the first period is half of this value:

$$\frac{1}{2}(\Delta^+ - \Delta^0)\delta S = \frac{1}{2}\frac{(\Delta^+ - \Delta^0)}{\delta S}(\delta S)^2$$

which is proportional to $(\delta S)^2 > 0$. The constant of the proportion is:

$$\Gamma := \frac{1}{2}\frac{(\Delta^+ - \Delta^0)}{\delta S} \ ,$$

i.e. the change of the Delta over time - the Gamma. The dynamic adjustment of the Delta over time is given by the Gamma for a long position.

This shows that gains are generated over time if the market maker Delta hedges his long position. Is this an arbitrage possibility, since the market maker with zero initial investment only needs to wait until the oscillations spontaneously generate cash. No. There are more factors in the transaction which we did not consider so far.

- Funding. If time elapses, the market maker has to pay back the funding plus interest rates. The interest rate costs are for a period δ with simple compounding equal to

  $$rC(T-t)$$

  with r a constant interest rate.

- Time decay. If time passes and all other parameters for the option price are fixed, then the option loses value. The value of the loss is

  $$\Theta(T-t)$$

  proportional to the period length $T-t$. The constant is the Theta Θ.

- Cash. The short position generates in each period an earning

  $$rS\Delta(T-t) \ .$$

  For simplicity we assume that this earning rate is equal to the funding rate.

The net cash gains/losses of the Delta hedged position are

$$\frac{1}{2}\Gamma(\delta S)^2 + r\Delta S(T-t) - rC(T-t).$$

All this cash flows have to be equal to the time decay, else arbitrage is possible, i.e.

$$\frac{1}{2}\Gamma(\delta S)^2 + r\Delta S(T-t) - rC(T-t) = -\Theta(T-t).$$

In the specific Black and Scholes model the dynamics of the underlying asset leads to

$$\Gamma(\delta S)^2 = \Gamma S^2\sigma^2(T-t)$$

which implies

$$\frac{1}{2}\Gamma S^2\sigma^2(T-t) + r\Delta S(T-t) - rC(T-t) = -\Theta(T-t)$$

or:

$$\sigma^2 S^2\Gamma + rS\Delta - rC = -\Theta\ . \tag{2.89}$$

In terms of derivatives this equation is a partial differential equation (PDE) in the unknown function C reflecting the no arbitrage free option price:

$$\sigma^2 S^2\frac{\partial^2 C(t,S)}{\partial S^2} + rS\frac{\partial C(t,S)}{\partial S} - rC = -\frac{\partial C(t,S)}{\partial t^2}\ . \tag{2.90}$$

We note that the drift does not matter but the risk free rate r like in the stochastic approach using martingales. Often it is not possible to solve analytically PDEs. Fortunately, this fundamental pricing equation is of the well-known type of diffusion equations. The solution exists - i.e. plugging the Black and Scholes formula (2.55) into the PDE with the boundary conditions shows that the pricing formula is a solution.

> The equation (2.89) together with the terminal condition for a call
>
> $$C(T,X) = \max(S-K,0)$$
>
> is called the Black-Scholes equation. This is the fundamental pricing equation.

The terminal condition reflects the contract under consideration, solving the PDE w.r.t. C gives its no arbitrage price. This equation is the equity market equivalent to the term structure equation for interest rate options.

Although we know the solution of the PDE of Black and Scholes one might wonder how one can solve such type of equations. There are indeed analytical methods which one can use to solve the equation. There is a deep mathematical result of Feynman and Kac which states that one can solve some PDEs (such as the Black and Scholes one) by running a Brownian motion, i.e. be calculating an expected value of a function of the Brownian motion (i.e. the martingale approach (2.88).

2.6.3 P&L - a Formal Approach

We start with an intuitive approach. The daily profit and loss **(P&L)** can be decomposed as a linear superposition of the Greeks contributions, i.e.

$$\begin{aligned}\text{Daily P\&L} &= \text{Delta P\&L} + \text{Gamma P\&L} + \text{Theta P\&L} \\ &+ \text{Vega P\&L} + \text{Rho P\&L} + \text{Mu P\&L} + \text{Higher order}\end{aligned} \tag{2.91}$$

where the sensitivities are in this order w.r.t. to the underlying, the change in Delta, time, volatility, interest rates, dividends and higher order derivatives. If we assume that implied volatility remains constant through time, interest rates are zero, dividends are zero and the higher order derivatives are negligible the daily profit and loss becomes

$$\text{Daily P\&L} = \text{Delta P\&L} + \text{Gamma P\&L} + \text{Theta P\&L} \tag{2.92}$$

$$= \Delta S + \frac{1}{2}\Gamma(\Delta S)^2 + \Theta\Delta t\ . \tag{2.93}$$

If the trader is **Delta-hedged** this reduces to

$$\text{Daily P\&L} = \frac{1}{2}\Gamma(\Delta S)^2 + \Theta\Delta t\ . \tag{2.94}$$

Theta and Gamma are related by

$$\Theta = -\frac{1}{2}\Gamma S^2\sigma_{im}^2$$

with σ_{im} the implied volatility, i.e. the parameter value which one has to put into the theoretical Black and Scholes option price formula such that the theoretical prices matches the market price of the option. The last equality follows from the Black and Scholes pricing equation with zero interest rates and Delta hedged positions, we finally get for the daily profit and loss

$$\begin{aligned}\text{Daily P\&L} &= \frac{1}{2}\Gamma(\Delta S)^2 - \frac{1}{2}\Gamma S^2\sigma_{im}^2\Delta t \\ &= \frac{1}{2}\Gamma\left((\Delta S)^2 - S^2\sigma_{im}^2\Delta t\right) \\ &= \frac{1}{2}\Gamma S^2\left((\Delta S/S^2)^2 - \sigma_{im}^2\Delta t\right)\ ,\end{aligned} \tag{2.95}$$

i.e. the final profit and loss is proportional to the difference of a floating leg (the stock variance) and a fixed leg (the implied volatility part). Therefore, for a Delta-hedged position the daily profit and loss is proportional to the squared percent change in spot price or equivalent the realized return which is the daily realized variance and the daily implied volatility. ΓS^2 is the dollar Gamma, i.e. the second-order change in the option price for a squared percent change in the spot price. Thus a trader who is long dollar Gamma will make money if the daily realized volatility is higher than implied, break even if they are the same, and lose money if realized is below implied. If we sum the daily profit and loss over time we obtain

$$\text{P\&L} \quad = \quad \frac{1}{2}\sum_{t=1}^{T}\Gamma_t S_t^2\left((\Delta S_t/S_t^2)^2 - \sigma_{im}^2\Delta t\right) \ . \tag{2.96}$$

The final profit and loss is a function of the daily volatility spread weighted by the Dollar Gamma, i.e. it is a path-dependent quantity. The formula shows that periods where Dollar Gamma is high will dominate in terms of P&L.

If we assume that S satisfies a geometric Brownian motion dynamics, the instantaneous variance $(\Delta S_t/S_t^2)^2$ equals

$$(\Delta S_t/S_t^2)^2 = \sigma^2\Delta t$$

with σ the **realized variance.** Therefore,

$$\text{Daily P\&L} = \frac{1}{2}\Delta t\Gamma S^2\left(\sigma - \sigma_{im}^2\right) \ .$$

Suppose the trader sells a call or a put, i.e he is Γ-short, i.e. $\Gamma < 0$. If he sold the option for $\sigma_{imp} < \sigma$, then he will make a loss; but if $\sigma_{imp} > \sigma$, then he will make a gain. The exact amount of the gain/loss depends on the stock price path (the integral); but the value is proportional to the difference between the volatilities used for pricing and the true volatilities. The result is intuitive and looks simple. But it is not. Suppose that a trading book has say 50 underlyings. Then, the full covariance matrix is of the size 50×50. Hence, the trader needs to have a view on 2450/2 pairwise correlations ρ_{ij}, 50 volatilities and their potential changes. This is not trivial. So far, we considered a single asset. Assuming that there are N assets which satisfy a system of coupled geometric Brownian motions, the result immediately extends to:

$$\text{Daily P\&L} = \frac{1}{2}\sum_{i,j}\left(\rho^{ij}\sigma^i\sigma^j - \rho_{im}^{ij}\sigma_{im}^i\sigma_{im}^j\right)S^iS^j\frac{\partial^2 C}{\partial S^i\partial S^j} \ . \tag{2.97}$$

Hence, besides volatility figures also implied and realized correlation between the assets matter. Equation (2.97) is the P&L of a trader who uses Black-Scholes and is Δ-neutral.

2.6.4 P&L and Risk Management, a General Approach

Chapter 3

Investment

3.1 Investment

3.1.1 Overview

Investment is about a choice to distribute \$ 1 today such that tomorrow $1+x$ follows with the goal $x > 0$. Since x is not known with certainty, unless it is a risk less investment, there is a risk that $x < 0$. Several questions arise:

- **Who** decides? What are the **objectives**, **constraints** and **opportunities**?
- What is the **method** in decision making? Economic analysis, rule based mechanism, replication of markets, views, optimal portfolio strategies or a mixture?
- How should we decide between the **risk and return trade-off**, i.e. what is the **objective**? Some questions related to this one are:
  - How **many** assets $x_1, x_2, \ldots, x_n$ should one choose?
  - How much **weight** should we give to each chosen asset at each date (portfolio weights)?
- What defines the **opportunity set**? What are the constraints, opportunities for the investment?
- Are the alternative asset combinations which deliver the same risk and returns but at lower total **costs**, i.e. the costs to buy the assets, the costs to sell the assets and taxation costs?
- How **liquid** are the assets: Can we sell them at any time as a price taker?

There is not a single, accepted approach how one should address these investment questions. Figure **??** illustrates a possible characterization of different methods. Three fundamental methods of investment are shown:

- Trust based or discretionary;

- Payoff based;
- Algotrading.

Trust based or **discretionary**, i.e. where the investor delegates the investment decision to portfolio managers either with the goal to reproduce the market or to beat the market. Broadly, this is the asset management approach. Originally, investments were linear, i.e. long and short positions but non-linear investments (options) were not used. Three types of objectives are common:

- Relative return. This approach defines the success of investing relative to a benchmark. A benchmark can be an index, a basket of assets or a portfolio which is the result of an optimization problem in (**Modern Portfolio Theory**). There are different ways how information is obtained which leads to portfolios different than the benchmark one. The Tactical Asset Allocation (TAA), i.e. the forecast of expected future returns conditional on actual information, and the Strategic Asset Allocation (SAA), i.e. the forecast of expected future returns, are two common types of information processing to forecast future returns. By construction, the TAA is more volatile than the long-term oriented SAA. Three factors impact the outcome of the return estimate: The information used in the conditional expectation for the TAA, the risk model and the degree of rationality to build the forecast. The information set can contain past price information of the assets but also macro economic information. The risk model, i.e. the assumed dynamics of the returns, is a second key ingredient. The deep and broad methods of econometrics are then used for the risk model and the information process to generate forecasts. Since not all factors can be included in such forecasts one often use these estimates together with subjective judgements to correct them. At this stage emotions, biases and other behavioral aspects matter. If one believes that one can systematically beat a benchmark one assumes that markets are not efficient. Put it different, if it is impossible to make profit by trading on the basis of the information used to predict asset returns one should on average not be able to generate a return above a benchmark return.
- Absolute return.
- Market replication. The objective is to obtain a broad, cheap and close replication of a market. This is often achieved using Exchange Traded Funds (ETFs).

Payoff based, i.e. the bank enters in a binding liability. This requires a trading unit, the willingness of the bank to take risks. Payoff based investments can be further classified considering the protection level and the investment horizon. In the first dimension, one distinguishes between capital protected products and products where capital is at risk. We stress that capital protection in payoff based investment is contingent on the issuer not defaulting (Lehman Brothers). The choice of the time horizon differentiates between opportunistic investments due to market disruptions, overreaction and longer term investments where not events but fundamentals provide the rationale. Payoff-based investment can be described in three dimensions:

- Underlying value.
- Payoff.
- Wrapping.

The dimension underlying value consists of all asset classes such as Equity (EQ), Fixed Income (FI), Commodity, Currency (FX), real estate, private equity. We consider Hedge Funds not as an asset class but as an investment strategy. This dimension has a second attribute which distinguish between single, basket or index investments. The payoff dimension can be decomposed in the following categories: payoffs with the possibility of discretionary interventions, i.e. a mixture or discretionary and payoff type investments and those where the payoff is given ex-ante as a mathematical expression. Examples are options and retail structured products. The third dimension **wrapping** is related to the costs of producing the investments, liquidity issues and taxation. For example to participate in a basket of stocks can be realized using an exchange traded fund (ETF) or a tracker certificate (RSP) or by simply buying all stocks on an individual basis. First, transaction costs are not the same. Second, taxation of the products with the same economic rationale can be different. Third, liquidity is not the same. Fourth, issuer risk the investor faces is very different. Fifth, time to market to produce the product are different. Sixth, life cycle management has a different complexity for the different solutions. Seventh, whether or not dividends of the underlying share are reinvested can be different with different impact on taxation.

Algotrading is a different way to generate earning based on market inefficiencies. Originally Algotrading ment to exploit arbitrage, i.e. to detect on the markets whether mispricings exist. For example, European put and call options satisfy a model independent equality (the Put Call Parity). This equation states that a call is a put and vice versa. Therefore, given the price of a put the price of the call follows. If one observes a different price arbitrage is possible: One designs a strategy where one makes a risk less gain in risky environment. Algotrading are computer programs which screen the market and search for such opportunities. Today algotrading strategies exist in many different forms. The above example belongs to so-called high-frequency trading, i.e. computers make elaborate decisions to initiate orders based on information that is received electronically, before human traders are capable of processing the information they observe. Other aspects of algotrading are not considering to exploit arbitrage opportunities but the algorithm serve other purposes such as for example to sell a large amount of stocks in an intelligent way where there is almost no price effect. A final use of algotrading is 'steal' flows. Consider a market maker with bid ask 100-100.5. Then a very fast machine observes this and quotes 100.1-100.4, i.e. the flow is taken away from the market maker. This kind of trading leads to exploding IT costs since it is a battle of technology. This kind of trading is so far not well understood from a systemic risk perspective. But given the almost zero value such a technological war adds one might wonder why regulators do not stop this kind of trading. The studies for the Government Office of Science of UK

(2011) state as key findings:[1]

- *Economic research thus far provides no direct evidence that high frequency computer based trading has increased volatility. However, in specific circumstances, self-reinforcing feedback loops within well-intentioned management and control processes can amplify internal risks and lead to undesired interactions and outcomes.*
- *Overall, liquidity has improved, transaction costs are lower, and market efficiency has not been harmed by computerised trading in regular market conditions.*
- *The nature of market making has changed, shifting from designated providers to opportunistic traders.*
- *Computer –driven portfolio rebalancing and deterministic algorithms create predictability in order flows. This allows greater market efficiency, but also new forms of market manipulation.*
- *Technological advances in extracting news will generate more demand for high frequency trading, while increased participation in this will limit its profitability.*
- *Today's markets involve human traders interacting with large numbers of robot trading systems, yet there is very little scientific understanding of how such markets can behave.*
- *Future trading robots will be able to adapt and learn with little human involvement in their design. Far fewer human traders will be needed in the major financial markets of the future.*

Despite the differences there is a **convergence** in the instruments used in the trust and the payoff based approach: Both approaches use linear (shares, bonds, futures, swaps, rates) and non-linear instruments (options, structured products).

The discretionary and payoff approaches can be defined more formal. Let $W(=1+x)$ be wealth at a future date or absolute return, B be a benchmark return. Then the asset only and the asset liability case objectives can be stated as follow:

- Asset only: Absolute return W, relative return $W-B$, replication B.
- ALM: With $S=A-L$ the surplus, we have absolute return S, relative return $S-B$ and replication B.

From an economic perspective one should write more general preferences than linear ones. Besides objective the constraint or opportunity set A has the following form:

- A depends on the wealth evolution (budget constraint) and on the decisions (portfolios).

[1] The authors of the three studies in the report are C. Furse, A. Haldane, C. Goodhart, D. Cliff, J.-P. Zigrand, K. Houstoun. O. Linton. P. Bond.

- A is a union of different constraints/opportunities: A risk constraint, an investment constraint (not more than 10 percent US stocks), an market admission constraints, a transaction cost bound, an incentive constraint are examples.
- It is obvious that adding a constraint reduces the return possibilities.

Finally, if ϕ is the investment strategy at a given date, wealth W' at a future date is given by:

$$W' = W + \phi \Delta S$$

with ΔS the price change of the investment (share, commodity, etc.) in a period. The strategy ϕ can be selected by a rule, discretionary or optimal as the solution of an optimization problem. Given this setup, discretionary investment means the solution of the problem

$$\max_{\phi \in A} W(\phi)$$

where the portfolio manager chooses ϕ, the chief investment officer defines the opportunity set A and the investor likes the whole optimization program.[2] In the payoff case, let $f(S)$ be a (non-linear) function of an underlying value S. Then $f(S)$ satisfies the constraints of the investor ($f(S) \in A$) and is optimal for the investor, i.e. $f(S)$ maximizes the objective. This shows that in the payoff case (i) the investor solves the optimization problem and not the portfolio manager. Given the optimal choice $f(S)$, the investor pays 1 USD today in exchange for $f(S)$ at a future date.

3.1.2 Diversification, Arbitrage Pricing Theory, Alpha, Beta

What do **Alpha** and **Beta** mean? Active management is often expressed using Alpha and Beta is used for passive investments. Alpha then denotes today each return which is attributable to active management and is above the risk less rate is. This view is not helpful for practice and has no theoretical foundation. The two Greek letters are **not** related to the trading frequency but to different types of **risk factors**. They are identifiable within to modern portfolio theory: Theories where the portfolio return follows to be a linear combination of factors. To understand this, we first consider the notion of diversification and then the decomposition of a return in risk factors.

Definition 3.1.1. *Systematic risk (aggregate risk, market risk, or undiversifiable risk) is vulnerability to events which affect broad market returns. Idiosyncratic risk (specific risk, unsystematic risk, residual risk, or diversifiable risk) is risk to which only specific agents/firms/assets are vulnerable. Idiosyncratic risk can be diversified away.*

'diversified away' means that the risk of a portfolio is reduced by adding more and more assets to the portfolio. We discuss the notion of **diversification**. We start with **abstract pure asset diversification**. There is no other type of diversification across

[2]Since wealth and other variables are random, one considers the expected value in the objective.

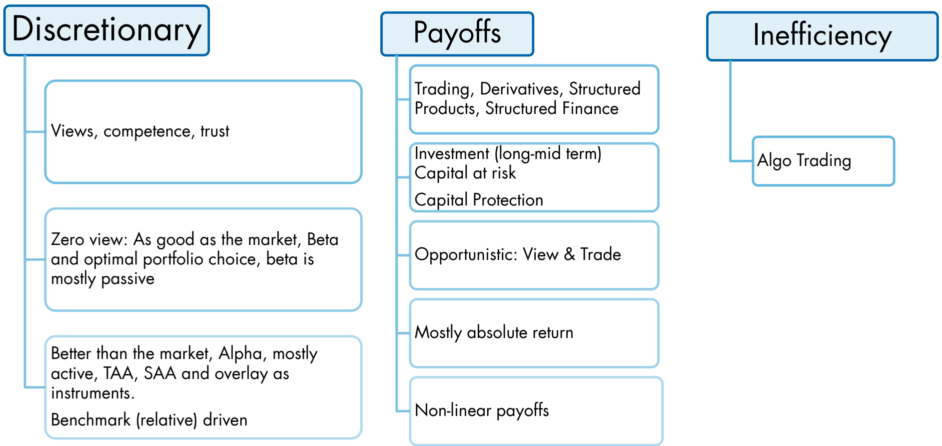


Figure 3.1: Different approaches to investments.

trades, models, information source, ideas and information processing. Assets are modelled as random variables. The basic result is:

Proposition 3.1.2. *Assume that returns are not correlated and investment is equally distributed. Then increasing the number of assets reduces portfolio risk monotonically.*

To prove the result, we first assume that there are N returns $R_1, \ldots, R_N$ of some assets which are **uncorrelated** and that there exists a uniform upper bound c for risk measured by the variance, i.e. $var(R_i) \leq c$ for all i. Using elementary properties of the variance,

$$var(\sum_{i=1}^{N} \phi_i R_i) = \sum_{i=1}^{N} \phi_i^2 var(R_i) \leq c \sum_{i=1}^{N} \phi_i^2$$

follows for an investment strategy vector ϕ. Supposing that the investment is equally-distributed, i.e. $\phi = \frac{1}{N}$, then

$$var(\sum_{i=1}^{N} \phi_i R_i) \leq c \sum_{i=1}^{N} \phi_i^2 = \frac{c}{N} \to 0 \ , \ (N \to \infty).$$

To eliminate portfolio risk completely in an equally-distributed portfolio with uncorrelated returns one has to increase the number of assets in the portfolio. Since volatility is the square of variance, adding new assets reduces volatility only as $\frac{1}{\sqrt{N}}$. But most assets traded in stock exchanges are **(positively) correlated**. Can we still eliminate risk completely?

Proposition 3.1.3. *Consider an equally-distributed portfolio strategy $\phi_i = \frac{1}{N}$ and N assets. Then the portfolio variance σ_p^2 is given by:*

$$\sigma_p^2 = \langle \phi, V\phi \rangle = \frac{1}{N}\overline{\sigma}_N^2 + (1 - \frac{1}{N})\overline{cov}_N$$

with the mean variance $\overline{\sigma}_N^2$ and the mean covariance $\overline{cov}_N$. If $N \to \infty$ and if the unsystematic risk $\frac{1}{N}\overline{\sigma}_N^2$ is uniformly bounded, then the systematic risk $\overline{cov}_N$ cannot be diversified away.

The proof is given in Appendix 7.7. The residual risk $\overline{\text{cov}}_\infty$, i.e. the average covariance, cannot be fully diversified and is called the **systematic (market) risk.** The diversifiable risk $\frac{1}{N}\overline{\sigma}_N^2$, i.e the average variance, vanishes as $N \to \infty$ and it is called the **unsystematic risk** of the portfolio. Figure 3.2 shows diversification for world wide stock market indices. Diversification allows to eliminate unsystematic risk or idiosyncratic risk. Market or systematic risk can only be reduced to the level of the average portfolio covariance. We often encounter the case that assets can be pooled. For example assets within an economic sector show a more similar dependence among them compared to assets in other sectors. To see that the above analysis also applies to the case where we have a macro structure on the assets we assume that the number of assets N can be decomposed in a set of N_1 assets and a second set $N_2 = N - N_1$ representing a second sector. To consider the limit N to infinity we assume that $N_1/N = c$ is constant. This implies that $N_2/N = 1 - c$ is also constant if $N, N_2 \to \infty$. If one repeats the calculation as in the proof of the last proposition for this two-sector economy one obtains in the infinite asset limit:

$$\sigma_p^2 = c^2\overline{\text{cov}}_1 + (1-c)^2\overline{\text{cov}}_2 + c(1-c)(c(1-c) - 1)\overline{\text{cov}}_{1,2}$$

with $\overline{\text{cov}}_i$ the mean covariance in sector i. This result shows that total portfolio risk is bounded again by the mean covariances, i.e. this risk cannot be diversified away. The formula shows that this non-diversifiable asymptotic risk arises from the two sectors and the cross-sector correlation. Compare this two sector case with the one sector one. For $c = 0, 1$ we are in the respective one sector boundary case. Choosing $c = 1/2$ it follows that portfolio variance σ_p^2 is lower than each of the two boundary values $\overline{\text{cov}}_1, \overline{\text{cov}}_2$. This shows that risk is not only reduced within an asset class by increasing the number of assets but also by increasing the number of different asset classes.

The figure shows that diversification within the global equity class alone is a **poor** concept - the different diversified regional markets are moving in the same directions. Only the extent of the yearly return variation is different:If one regional market is positively performing almost all others behave in the same way. Hence, instead of investing in many local markets one can equally invest in a single index representing the world wide stock markets. For a true diversification the figures shows that other asset classes

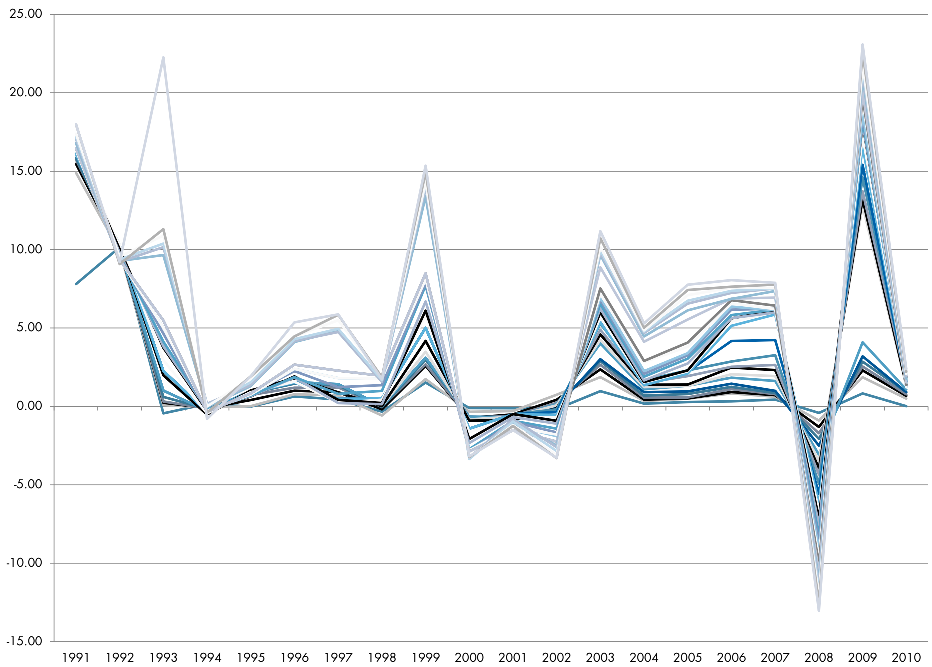


Figure 3.2: Diversification for world wide stock market indices. On the vertical axis the annual linear returns are shown. The following stock market (indices) are shown: American SE, BOVESPA, Buenos Aires SE, NASDAQ OMX, NYSE Euronext (US), TSX Group, Australian SE, Bombay SE, Bursa Malaysia, Hong Kong Exchanges, Korea Exchange, National Stock Exchange India, Shanghai SE, Shenzhen SE, Singapore Exchange, Taiwan SE Corp, Tokyo SE, Athens, BME Spanish Exchanges Madrid, Borsa Italiana, Budapest SE, Deutsche Börse, Egyptian Exchange, Istanbul SE, Johannesburg SE London SE, SIX Swiss Exchange, Tel-Aviv SE. *Source: World Federation of Exchanges, 2012.*

are needed.

So far diversification was across **assets**. But there is also diversification across **methods**. Diversification across methods means a reduction of model risk: The risk that each single approach in investment faces the risk of being misspecified. Different methods:

- Information sources. What are the sources of information which are used in decision making? Fundamental values, technical information, market information (flows), positions, macro information.
- Strategies across regions and assets classes, i.e. invest/exit and long/short. Focus not diversification.
- Quantitative models such as market signalling models (momentum signals, VaR measures, information ratios, spread information).
- Trades. Diversification across different trades. For example in FI and FX duration, cross-market, government, curve, credit and currency.

This characterization is used by JP Morgan (2004). Diversification can therefore be considered as:

$$\text{Diversification Assets} \times \text{Diversification Methods} \to \text{Trades} .$$

This wider interpretation of pure asset diversification has also drawbacks. First, different methods can be conflicting: What a strategist propose can be conflicting with duration risk. This requires a procedure to resolve conflicts or to make a decision between conflicting alternatives.

Is there a foundation for this double diversification approach? One might argue that more diversification is always better. But this is in general not true. First, if one has the analytical skills to understand deep economic relations and if one has the capacity to transform the view into a trade then one is not interested in diversification per se. Second, portfolio theory assumes that probabilities of future outcomes are known. But there is ambiguity in these probabilities. Is there relationship between risk and ambiguity? Izhakian (2011) shows that in most cases ambiguity cannot be diversified without increasing risk. This implies that with ambiguity it is often not optimal to hold a fully diversified portfolio. Assuming that investors are averse both to risk and ambiguity, they would like to minimize both of them. In the modelling approach[3] not only returns are random but the probabilities of these returns are themselves random. In this model, the degree of risk is measured by the variance of returns (as usual) and the degree of ambiguity can be measured by the variance of the probability of loss or gain. Working within the family of normal distributions random probabilities means random means and random variances of the distributed asset returns. The degree of ambiguity is a matter

[3] Izhakian calls the model expected utility with random probabilities (EURP).

of the classification of returns as losses or gains, relative to a reference point. Izhakian proves that ambiguity and risk are inversely related under quite weak assumptions: Under these assumptions ambiguity cannot be diversified without increasing risk. Formally, the intuition is as folows. Let X be a random variable. The expected value of X is then a double sum - a sum over the probability weighted outcomes (risk) and a sum over all probabilities which one considers (ambiguity). The random mean is then a random variable where one takes only the sum over the weighted outcomes. The central concept is that the probabilities of outcomes are random; thus, as the degree of risk can be measured by the **variance of outcomes**, the degree of ambiguity can be measured by the **variance of probabilities**. Then the sum over all probabilities where the random variable X is lower than a reference point is the expected random loss. To be more concrete, we consider an example.

- Assume that the probabilities of d and u are known: $P(d) = P(u) = 0.5$.
- The expected return is 5 percent and standard deviation is 15 percent. Ambiguity is zero.
- To introduce ambiguity, $P_1(d) = 0.4$ and $P_1(u) = 0.6$ and $P_2(d) = 0.6$ and $P_2(u) = 0.4$ are the uncertain probabilities. The new distribution which describes the probability that either the first or second randomness is realized is an equal distribution: With 50 percent the first probability P_1 (P_2) is drawn.
- Assume that a negative return is considered a loss. Ambiguity, which is four times the variance of the probability of loss, is then 0.2.

The common thread between risk and ambiguity is the random variance. The higher random variance is, the higher is risk. The impact on ambiguity is the opposite. Since a higher random variance means a flatter random probability density function, a lower ambiguity follows: Formally, if the reference point satisfies some conditions:

$$\frac{\text{Ambiguity}}{\text{Random Variance}} \leq 0 \ .$$

Hence, adding an asset increases its ambiguity as it decreases its risk. Can ambiguity be diversified away in an asset portfolio or are two assets less ambiguous when combined than each asset separately? Again the result is in the opposite direction than for the risk dimension. For risks, an incresing correlation means increasing risk. But if correlations of the asset return increases the variance of the returns is also increasing. But as we stated above an increasing random variance reduces ambiguity. If an investor considers risk and ambiguity, holding a fully diversified portfolio is not optimal sinc since for a given expected return, minimizing risk increases ambiguity.[4]

[4] The paper considers some puzzles: Individual investors tend to hold very small portfolios—only 3-4 stocks, i.e., underdiversification (Goetzmann and Kumar (2008)), or they choose not to participate in the stock market (Guo (2004) and Bogan (2008)). Expected volatility is higher than realized volatility (volatility risk premium Eraker (2004), Car and Wu (2009) and Drechsler (2012)).

Consider a bank which uses a modern portfolio theory model. Such a model faces return and volatility risk, i.e. risk that the estimates of the past which are used to calibrate the model are ambiuous. To consider ambiguity the bank can ad hoc change the fully diversified portfolio outcome of the model or they can use other sources in investment decision making - analysts, strategiy analyists, trade information. In this sense double diversification as discussed above can be conidered as method to control risk and ambiguity in investment decision making.

We consider **Arbitrage Pricing Theory (APT)**. We follow Wang (2003). APT was developed primarily by Ross (1976a, 1976b). It is a one-period model in which every investor believes that the stochastic properties of returns of capital assets are consistent with a factor structure: It starts with specific assumptions on the distribution of asset returns and relies on approximate arbitrage arguments. Ross argues that if equilibrium prices offer no arbitrage opportunities over static portfolios of the assets, then the expected returns on the assets are approximately linearly related to the factor loadings. APT assumes that asset returns R are driven by N assets:

$$R = \mu + \beta f + \epsilon \tag{3.1}$$

with β a $N \times K$ matrix, f the K factors, ϵ is the risky asset's idiosyncratic random shock, f, ϵ are both assumed without loss of generality to have mean zero, idiosyncratic ϵ risk is assumed to be to be uncorrelated across assets and uncorrelated with the factors and μ is the expected asset return. Given the assumption, μ equals the expected return $E[R]$. Each factor f_j is written as the 'surprise':

$$f_j = F_j - E[F_j] \ .$$

β_{ik} gives the sensitivity of return i with respect to news on the k-th factor. It is called the factor loading of asset i on factor f_k. Using a no arbitrage argument one derives:

Proposition 3.1.4 (APT). *Assume the factor representation (3.1) holds with all assumptions. Then the expected return μ_i for any asset i depends only on its factor exposure*

$$\mu_i \simeq \mu_f + \sum_{k=1}^{K} \beta_{ik}(\overline{R}_{f_k} - \mu_f) \tag{3.2}$$

with $\overline{R}_{f_k} - \mu_f$ the premium on factor k and μ_f the risk free rate.

This model gives a reasonable description of return and risk. Contrary to the CAPM, which can be seen as a particular case of the APT, the is no need to measure the market portfolio. The model does not says what the right factors are which also can change over time. Conceptually, APT is based on approximate no arbitrage and the factor models of returns. CAPM, which we below after the Markowitz model, is based on investor's portfolio demand and equilibrium arguments. Since (approximate) arbitrage is a more robust assumption, APT is often preferred in practice to CAPM. Some immediate consequences of the Proposition are:

- Any well-diversified portfolio is exposed only to factor risks, i.e. idiosyncratic risk can be diversified away.
- A diversified portfolio that is not exposed to any factor risk must offer the risk-free rate.
- There always exist portfolios that are exposed only to the risk of a single factor. That is, given say two portfolios which are exposed to two factors one can choose a convex combination of the two portfolios such that resulting portfolio is only exposed to one risk factor.

Instead of discussing the not difficult proof, we prefer to consider examples to gain intuition. Consider two factors f_1, f_2. The first one describes unexpected market return, the second one unexpected inflation. We assume a risk less rate of 5 percent a premium on market return of $8\% = \overline{R}_{f_1} - r_f = 13\% - 5\%$ and a premium on inflation of $-2\% = \overline{R}_{f_2} - r_f = 3\% - 5\%$. We assume that there is only factor risk, i.e. idiosyncratic risk is zero. The APT equation 3.2 reads for an asset X:

$$\begin{aligned} \mu_X &= \mu_f + b_1(\overline{R}_{f_1} - r_f) + b_2(\overline{R}_{f_2} - r_f) \\ &= 0.05 + b_1 \times 0.08 - b_2 \times 0.02 \ . \end{aligned}$$

Consider the specific asset X where both Betas are 1.0. Then, $\mu_X = 11\%$ follows. Suppose that the expected return of X is different, say 10%. Then there is arbitrage. To see this, we invest USD 100 twice in the two parts of the above portfolio (the so-called factor portfolios): We invest the amount in the portfolio which pays the return $0.05 + 0.08$ and the same amount with return $0.05 - 0.02$. We sell the 100 USD of low priced portfolio with the expected return μ_X. In summary we are so far long the two factor portfolios and short the X-portfolio. To prove that there is an arbitrage, we should state with initial investment of zero. So far we have USD $200 - 100$ total investment. To get zero, we simply sell another USD 100 in the risk free asset. All four portfolios require zero initial investment, they bear no risk and they pay $1 = 13 + 3 - 10 - 5$ USD for sure. Hence, this would be an arbitrage. This shows that the APT must hold in absence of arbitrage if $b = 1$, i.e. if there are only factor risks. What if an asset also bears idiosyncratic risks? Then by a contradiction argument shows that if the APT equation is not true for most assets, again arbitrage is possible.

After the discussion of diversification and APT, we consider the notions of Alpha and Beta:

- Beta. This is the proportion of the yield, which is associated to one or more systematic risk factors (such as a stock index, a bond index, etc.).
- Alpha is the residual amount.

Thus, in this view, Alpha has nothing directly to do with active Management but with non systematic risk factors. As more and more risks are added over time, Alpha **decreases**

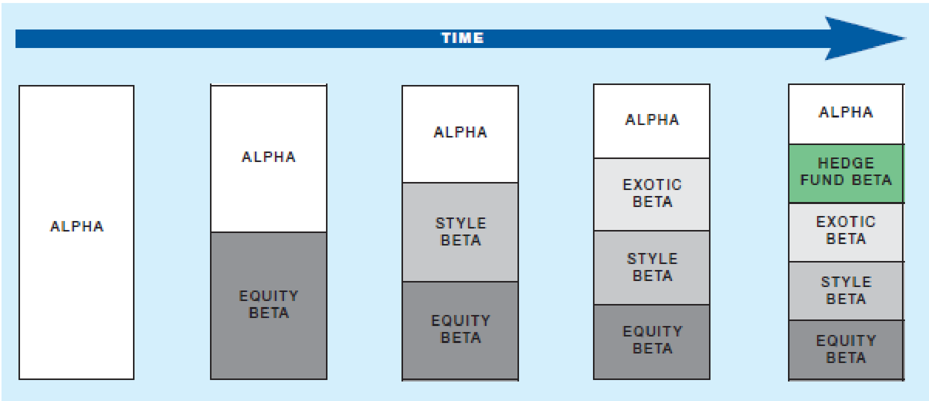


Figure 3.3: Development of the Alpha over time. *Source: Berger and Kalliber, AQR, 2008*

over time. In the historical development of each yield above the risk less rate was denoted Alpha, see Figure 3.3: In each portfolio with a return above the risk free rate this result was assigned to the portfolio manager. The use of indices reduced the Alpha for the first time. The managers were then able to express a large portion of the return with the commonly held market risk. The CAPM model provides, for example, the following decomposition for the expected Return μ_i of the title i:

$$\mu_i = \mu_0 + \beta_i(\mu^T - \mu_0) + \alpha \ ,$$

with $\beta_i = \frac{\text{Cov}(\mu_i,\mu_M}{\text{Var}(\mu_M)}$ and $\alpha = 0$ in the CAPM.

Some managers succeeded to achieve systematically returns above market returns. They followed simple rules on subsets of the broadly diversified indices. They, for example, overweight small-cap stocks and underweight large-cap stocks. Fama and French (1993) have shown in their work, using a three-factors model, that the returns of portfolio can be better explained if one adds to the CAPM distinctions of such market structures such as Large-cap stocks, small-cap stocks, growth stocks, etc. This explained a further part of the Alpha - the so-called **style risk premia** or the style Beta. Other empirical studies have shown that a further part of the returns can be assigned to alternative investments - the **alternative risk premia** or exotic Beta. The last part of the transformation of Alpha to the Beta is based on the hedge fund industry. Many hedge funds apply the same concepts. This results in that a further part of the Alpha is transformed

Asset Class	**Market Beta**	**Style Beta**	**Hedge Fund Beta**
Equities	Equity indices	Size, Value, Momentum	Merger Arbitrage
Fixed Income	FI market	Credit Spreads, HY Spread	Convertible Arbitrage
FX	FX market		Carry Trade

Table 3.1: Transformation of Beta for different asset classes.

to the Beta. In addition to the presented subdivision, there are other approaches. For example, the hedge fund Beta can be subdivided into a part called the inefficiency Beta. There are market situations in which you can exploit through systematic action structural imbalances. The imbalances can be generated by market participants, regulators, central banks or tax authorities.

The reduction of Alpha is for several reasons beneficial to investors: They are increasingly less exposed to non-identified risks and costs of portfolio management are reduced, since the investment in the Beta-parts is not specific to the asset manager's skills but can be achieved at lower costs using simple rules. Table 3.1 shows how for the asset classes of equities, fixed income and currencies the general market Beta, style Beta and Hedge Fund Beta is replicated.

The word 'arbitrage' comes from practice and should not be confused with the theoretical concept of no arbitrage. The various premia for e.g. 'value', 'size' or 'credit spread' should be both implementable for long and short positions. Figure 3.4 shows different implementations for the Betas of Hedge Funds and Styles, see Brian et al. (2009).

More important than the risk-return characteristics of the individual Beta components are the dependence structures, i.e. how do the individual risk premia correlate? The figure shows that many risk premia only slightly positive or even negative correlate with each other. Thus they offer real diversification to investors. Two points should be noted, however. The correlation between high yield and credit spreads is large. This is not surprising since both involve creditworthiness risk. Thus, one can dispense with one of the two factors if there is no significant basis risk. The second observation relates to the period of investigation, ie it stops at the outbreak of financial crisis. A fact of a crises is that the correlation structures can change dramatically - a negative or weak correlation can become strongly positive for example. Comparing these risk premiums with those of the market Betas, i.e. comparing the correlation between 'value' and the MSCI shows, that many of the style and Hedge Fund risk premia diversify. Others like 'Merger Arbitrage' or 'High Yield Spread' possess a strong positive correlation with stock indices - i.e. they are largely redundant.

3.1.3 Active vs. Passive Investments

There is an ongoing debate whether passive or active portfolio management pays for the investor. We start with two basic results. W. Sharp states that it must be the case that

Risk premium	Long Position	Short Position
Value	MSCI World Value	MSCI World Growth
Size	MSCI AC World Small Cap	MSCI AC World Large Cap
Momentum	World Momentum (simulated)	MSCI World
Credit Spread	Merrill Lynch US Corporate (AAA)	Merrill Lynch US Treasury
High Yield Spread	Merrill Lynch High Yield US Corporate	Merrill Lynch High Quality US Corporate
Term Spread	Merrill Lynch US Treasury 10+ years	Merrill Lynch US Treasury 1-3 years
Merger Arbitrage	Target	Acquirer
Conv Arbitrage	Convertible Bond	Underlying Stock
Currency Carry Trade	3 Highest Interest Rate G10	3 Lowest Interest Rate G10
Currency Value	3 Most Undervalued G10	3 Most Overvalued G10
Currency Momentum	3 Best Performing G10	3 Worst Performing G10

	Value	Size	Mom	Credit Spread	HY Spread	Term Spread	Merger Arb.	Conv. Arb.	Carry Trade	Curr. Value	Curr. Mom.
Value	1										
Size	0.11	1						< 0.25			
Momentum	-0.47	0.15	1					> 0.50			
Credit Spread	0.07	0.06	0.15	1							
High Yield Spread	0.06	0.26	0.20	0.56	1						
Term Spread	-0.05	-0.02	0.18	0.04	-0.09	1					
Merger Arbitrage	-0.02	0.18	0.33	0.29	0.49	-0.12	1				
Convertible Arbitrage	-0.02	0.18	0.52	0.25	0.39	0.17	0.44	1			
Carry Trade	0.00	0.10	0.35	0.41	0.45	0.08	0.37	0.58	1		
Currency Value	0.10	-0.04	-0.19	0.20	0.20	-0.03	0.00	-0.20	0.32	1	
Currency Momentum	-0.10	-0.20	-0.14	-0.13	-0.26	0.01	-0.20	-0.20	-0.01	-0.05	1

Figure 3.4: Implementations for the Betas of Hedge Funds and Style (left panel) and correlation of the style and hedge fund risk premia from May, 1995 to October, 2008 (right panel). *Source: Briand et al., 2009*

1. before costs, the return on the average actively managed dollar will equal the return on the average passively managed dollar and
2. after costs, the return on the average actively managed dollar will be less than the return on the average passively managed dollar.

He claims that every statement holds for any time period. The claims imply that for active management the selection process of the asset managers is key since without skills on average active management is worthless. A market must be chosen such as MSCI index for example. By definition a passive investor always holds a fraction of the market. An active investor is one who is not passive. His or her portfolio will differ from that of the passive managers at some or all times. Because such managers tend to trade fairly frequently, they are called 'active'.

Let R_M, R_P, R_A the market -, passive investment - and active investment returns, respectively. Over any specified time period, the market return will be a weighted average of the returns on the securities within the market, using beginning market values as weights, i.e.

$$R_M = \lambda R_P + (1-\lambda) R_A$$

with λ the weight factor. Since $R_M = R_P$ before costs, the above equation implies $R_M = R_A$. This proves claim 1.

To prove claim 2, we note that the costs of actively managing a given number of dollars exceeds those of passive management. Active managers must pay more for research and for more frequent trading. Because active and passive returns are equal before cost, and because active managers bear greater costs, it follows that the after-cost return from active management must be lower than that from passive management. This proves the second claim.

Claim 1 implies that in the group of active investors there are some which beat the market return and a second group which under perform. These results do not depend on any behavioral assumptions or advanced statistical concepts - they follow from basis arithmetics.

Using claim 1 one could conclude that active management in the aggregate can never outperform market performance. To see how active management can nevertheless outperform performance, we consider rates of return to capital for investments in China. Understanding how the Chinese markets are regulated it follows that the active performance should be split into a R_A^p private sector and a state sector R_A^s which is dominated by the private sector, see Figure 3.5. We have

$$R_M = \lambda R_P + \mu R_A^p + (1-\lambda-\mu) R_A^s$$

with λ, μ the weight factors. Although $R_M = R_P$ holds still true, the active returns are once higher and once lower than the market return. But how do we know that the

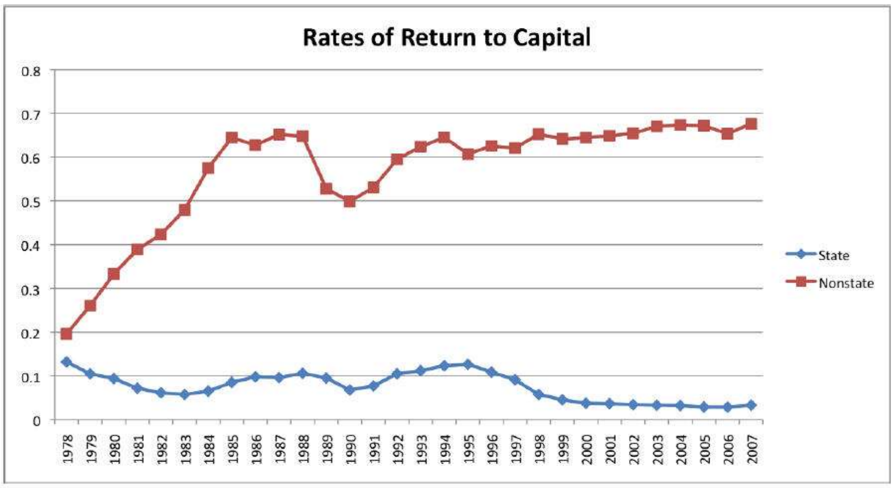


Figure 3.5: Rates of return to capital by sector for China. *Source: Xiaodong Zhu, University of Toronto.*

private sector outperforms the state owned one and do this last for all future dates? The necessary insight is that China's remarkable growth performance is driven by rapid productivity growth in the non-agricultural non-state sector. But there are financing constraints which lead to under investment in the non-state sector such as limited access to external financing by non-state firms. Furthermore there is high business risk, taxation is not favorable for non-state firms and costs of investment are high. Despite high domestic saving rates, many high return investment projects in the non-state sector are not financed. In summary there are structural causes for the difference in the rates of return before costs. Successful investment is not the art of understanding or estimating market signals **but** to understand underlying economic, legal or institutional structures. If an investor can find the possibility to invest in the non-state sector then even after costs the return of non-state investments dominates the state investments. This will not last for ever since improvements in capital markets will remove the imbalances. But for a short or midterm investor active investment into Chinese stocks 'dominates' passive investment - if he gets access to the Chinese' assets.

There is no guarantee that active investment in the above example dominates passive one. But the example shows how knowledge based active management can be more valuable than reasoning based on statistical estimates of past returns and risks.

Active management often has two components: A passive one which represents long-term goals in a benchmark portfolio and an active portfolio on short or medium-term which represents views. Active management defines deviation from the benchmark to benefit from market opportunities. In such a setup the passive portfolio stabilizes the whole investment about what one expects to be optimal in the long term. What defines the success of the active strategy? First, skills to identify opportunities matter. The second factor is the number of trades. This two factors **skill** and **activity** impact the success of active management as follows. Assume for simplicity that returns are normally distributed with mean zero and variance σ^2. The profitable trades are defined as having a positive return. The expected return of the profitable trades equals $\sigma\sqrt{\frac{2}{\pi}}$.[5] Consider two traders. One trader is always successful, the other one in x percentage of all trades. Both trade n times. The information ratio, i.e. the measure of the trader's generated value, is defined as the excess return over risk. The **Information Ratio (IR)** is a measure of the risk-adjusted return of a financial security (or asset or portfolio). It is defined as expected active return divided by tracking error, where active return is the difference between the return of the security and the return of a selected benchmark index, and tracking error is the standard deviation of the active return. The expected return is

$$\underbrace{xn\sigma\sqrt{\frac{2}{\pi}}}_{x\% \text{ success in } n \text{ trades}} - \underbrace{(1-x)n\sigma\sqrt{\frac{2}{\pi}}}_{1-x\% \text{ failures}} = (2x-1)n\sigma\sqrt{\frac{2}{\pi}} .$$

[5]The expected value $E[R\chi_{R\geq 0}]$ of successful trades is given by: $\frac{1}{\sqrt{2\pi}\sigma}\int_0^\infty xe^{-x^2/\sigma^2}dx = \sigma\sqrt{\frac{2}{\pi}} \sim 0.8\sigma$.

If $x = 1$, the trader is always correct. Since risk scales with the square-root of the number of trades (or time), risk equals $\sigma\sqrt{n}$. This gives the information ratio

$$\mathrm{IR} = (2x-1)\sqrt{\frac{2n}{\pi}} \ .$$

Therefore, for a fixed success rate x an increasing trading frequency increases the information ratio and raising the trading frequency encounters diminishing returns. Numerically, a IR of 0.5 needs a success rate x of two-third if the trader trades quarterly. We remark that the volatility does not enter the IR - this is particular to the assumption of a normal distribution. Assume that a single trade induces costs $k > 0$. The IR after costs then reads

$$\mathrm{IR}_{\mathrm{post}} = (2x-1)\sqrt{\frac{2n}{\pi}} - kn \ .$$

This simple extension shows that increasing the trading frequency no longer increase the IR but that there is an optimal level which depends on the ratio between the success rate and the cost factor.

3.1.4 Efficiency of Markets

Since Eugene Fama, many academics believe financial markets are too efficient to allow for repeatedly earning positive Alpha, unless by chance. To the contrary, empirical studies of mutual funds usually confirm managers' stock-picking talent, finding positive Alpha. However, they also show that after fees and expenses are deducted, the effective Alpha for investors is negative - as it must be according to the two claims of W. Sharpe. Nevertheless, Alpha is still widely used to evaluate mutual fund and portfolio manager performance, often in conjunction with the Sharpe ratio and the Treynor ratio.[6] This raises the question to which extend one **predict** returns?

Intuitively the time horizon for prediction matters. To predict a share price in one hour is most likely close to a random event. But to forecast the return over 2 years randomness is superimposed by non-random factors such as quality of the production, human capital, skills of the management, etc. Randomness is likely to enter more as external events, i.e. technological revolution or market disruptions for example. The question is to what extend are the portfolio managers able to identify and value the fundamental values of the firm.

The Efficient Market Hypothesis (**EMH**) considers market structures where the forecast of future returns is completely random or in other words, where it is impossible to make profit by trading on the basis of the information used to predict asset returns. If markets are efficient passive investments, i.e. investing in an index and holding the investment to maturity is the best strategy. This raises the debate between active or

[6]The Sharpe ratio is a measure of the excess return (or risk premium) per unit of deviation in an investment asset or a trading strategy and the Treynor ratio relates excess return over the risk-free rate to the additional risk taken; however, systematic risk is used instead of total risk.

passive investments. While a significant part of the academic literature gives evidence that active investment does not adds value significantly in a given set of assets under consideration in practice both approaches coexist simply because there markets which are not efficient and where an active strategy can make payoff attainable which are not in a passive investment. The test of efficiency is often related to **statistical efficiency** - to what extent are future returns predictable. But there are inefficiencies which are **not random** in nature. Market segmentation and fragmentation are two examples. I

How is the EMH defined? Samuelson (1965) stated that in an informationally efficient market price changes should be unpredictable. Fama (1970) then argued that a market is efficient if prices fully reflect all available information. After the work of Fama there have been several refinements and precisions in the statement of the EMH. Malkiel (1992): *'...A capital market is said to be efficient if it fully and correctly reflects all relevant information in determining security prices. Formally, the market is said to be efficient with respect to some information set.... if security prices would be unaffected by revealing that information to all participants. Moreover, efficiency with respect to an information set ... implies that it is impossible to make economic profits by trading on the basis of [that information]...'*

One consequence of the EMH is that unless an investor is just plain lucky, it is impossible to exploit the market to make an abnormal profit by using any information that the market already knows. Another consequence is that for someone without any such private information, it does not make any sense to talk about 'undervalued' or 'overvalued' individual securities, sectors or markets. Finally, it is impossible for any other portfolio than the total-market portfolio to have both a higher expected return and lower risk than the total-market portfolio, i.e. the total-market portfolio is always located on the 'efficient frontier' in academic risk/return models.

There are three definitions of the information sets which in turn lead to three EMHs:[7]

- The information set includes only the history of the prices or returns themselves -

[7]

- If agents are rational the three different forms of the EMH can be described as follows in forecasting prices S of assets. Let $\mathcal{F}_s$ be the information set generated by the price history of the asset S only starting at time 0 up to time s. Then for $t > s$

  $$E[S_t|\mathcal{F}_s]$$

  is the best guess of future prices S_t. It follows that the conditional expectation is the orthogonal projection on the information set $\mathcal{F}_s$. If $\mathcal{N}_s$ denotes the set of public information which is not part of the price information, the semi-weak form implies

  $$E[S_t|\mathcal{F}_s \cup \mathcal{N}_s] \ .$$

  Since the conditional expectation has no vector space structure on the information sets, such models are difficult to handle analytically and general, intuitive results can not be expected. A second obstacle is the difficulty to define the set of public information. Finally, if $\mathcal{P}_t$ denotes the **weak** form of the EMH follows.

- The information set includes all information known to all market participants (publicly available information) - the **semi-strong** form of the EMH follows.
- The information set includes all information known to any market participant (private information) - the **strong** form of the EMH follows.

The assumptions for market efficiency to hold true are strong. For example, it requires that equal perception of risk among all investors holds. But a T-Bill might be considered risk free for an US investor but not for a German one, the assumption of the same risk aversion is clearly wrong, some investors have an asset-only view others such pension funds an asset liability one with typically lower risk tolerance and investors also differ in their time horizon for the investment. This remarks make it plausible why there is active portfolio management.

How can one test the EMH? If markets are efficient abnormal profits should not be possibly by trading on information. This is an operational concept which allows for empirical tests. One can consider profits by professional market participants. If after adjusting for risk the returns are superior than the returns of the others, markets are not efficient. The problem is how to observe the information which the agents use. A second approach is to consider hypothetical trading strategies based on explicitly specified information set. Do they earn superior returns? To test for it one needs (i) to specify the information set and (ii) to specify the model for risk. We consider the second approach.

What about the model for risk: How do we define abnormal returns? One can define abnormal return as the difference between realized and expected returns. But to capture the expected return a risk model is needed (CAPM, APT). Then, we state the Null Hypothesis for the EMH:

$$E(\text{Abnormal Return}_{t+1}|\mathcal{F}_t) = 0$$

where $\mathcal{F}_t$ is one of the three information sets. If abnormal return is predictable using the information $\mathcal{F}_t$, the hypothesis of market efficiency is rejected. This shows that a test of EMH contains a **joint hypothesis (JHP)**: (i) That markets are efficient and that (ii) the right risk model is chosen. This implies that market efficiency can never be rejected. One cannot know if the market is efficient if one does not know if a model correctly stipulates the required rate of return. Consequently, a situation arises where

private information which the strong form formally implies

$$E[S_t|\mathcal{F}_s \cup \mathcal{N}_s \cup \mathcal{P}_s] \ .$$

We note that the weak form is the standard one used in financial engineering and mathematical finance. Furthermore, in the theory of asset pricing the expectation of the asset S discounted by an appropriate numeraire is a martingale, i.e. the best guess of the future discounted asset price given present information is the present value of the discounted asset price.

either the asset pricing model is incorrect or the market is inefficient, but one has no way of knowing which is the case. In other words there is interference between model risk of asset pricing models and the potential efficiency in markets which one cannot disentangle. The debate between 'Rational Finance' and 'Behavioral Finance' can be framed in terms of the JHP.

To test for the EMH a new taxonomy of tests is proposed (see Franzoni (2008) for details). First, there are tests for predictability, both cross-sectional and for time-series. These are tests the JHP supporting the weak from EMH. Second, on focus on event studies which represent test for the semi-strong EMH. One computes returns after the release of public information and sees if returns are different from zero after the events. Given the short period, risk adjustment is not important. These are very close tests of the corresponding EMH. Many events or news make stock markets react quickly: Prices go up quickly supporting the semi-strong hypothesis. But when investors become aware that the news should induce a short position instead of a long position, one observes often a delayed and incomplete correction downwards, i.e. a violation of the semi-strong hypothesis. Third, to test the strong EMH, one tests for private information: One focus on hedge fund or mutual fond performance for example.

The results of the tests of the tree types are roughly as follows. In the **short run** there is almost no predictability, i.e. the market seems to be adequately described by a random walk. If one switches to longer time horizons returns become **more and more predictable**. The JHP clearly manifests itself. In the long run returns become predictable using valuation ratios such as the earning-price ratio. Randomness is important in the short run but in the long run deterministic, possibly time-varying relations, determine the returns. For a more detailed and in depth analysis we refer to the literature.

The EMH does not tell how individuals behave given an information set. In rational general equilibrium models information is modelled as a stochastic process and individuals use actual information to forecast future prices using conditional expectations. They estimate future values by conditioning on the present information set. The conditional expectation is mathematically the best estimate of future prices if all agents in the economy behave in this way. There is strong evidence from psychology, experimental economics and field experiments that people fail to apply the laws of probability correctly. Framing the information set, deviations from rational forecasting are even stronger. By assuming that information is in the prices the EMH does not tell anything of how information is perceived, understood, valued and transformed into economic activities. Therefore if a hypothesis fails to be supported behavioral economists attribute the imperfections in financial markets to a combination of cognitive biases such as overconfidence, overreaction, representative bias, information bias, and various other predictable human errors in reasoning and information processing. The EMH does not require that investors behave rationally. When faced with new information, some investors may overreact and some may underreact. Markets would not behave the way they do in the real world if everyone always reacted in the same perfectly rational way to new information. All that is required by the EMH is that these overreactions and underreactions be random enough and cancel each other out such that the net effect on market prices cannot be exploited to make an abnormal profit. Stated different, irrationality is irrelevant as long as it is unpredictable and not exploitable. Even the entire market can behave irrationally for a long period of time and still be consistent with the EMH, again as long as this irrational behavior is not predictable or exploitable. Thus crashes, panics, bubbles, and depressions are all consistent with a belief that markets are efficient.

Andrew Lo (2010) and his collaborators recently tested the human ability to detect **randomness** from **non-random** financial returns. The logic of their paper is as follows. Pick a time series p of historical prices $p_0, p_1, \ldots, p_T$ with the returns $r_t = p_t - p_{t-1}$. Then generate a new series of prices p^* by permuting the historical prices. Such a random permutation p^* does not alter the marginal distribution of the returns but it does destroy the time-series structure of the original series, including any temporal patterns contained in the data. Hence, all moments of the two series p and p^* agree but the latter one has no time series pattern which can be used for predictions: One can test individual behavior in visual pattern recognition. The two price series are shown and the individual is asked to decide which of the two moving charts is the real one by clicking on it. The individual is informed immediately whether the choice was correct or not. For each data set, the user is shown approximately 35 pairs of moving charts and asked to make as many choices. The null hypothesis is that participants cannot distinguish between actual and randomly generated price series, i.e. their choices should be not better than random guesses. Testing the null hypothesis means to calculate the p-value of obtaining at least as many correct guesses when guessing at random. Formally, if s is the number of individuals, c the number of shown charts, X the number of 'heads' in $n = s \times c$ independent tosses of a fair coin and g the number of correct guesses, we have.

$$\text{p-value} = P(X \geq g) = \sum_{i=g}^{n} \binom{n}{i} / 2^n .$$

There is strong evidence against the null hypothesis since the p-values are at most 0.503 percent for each of the eight data sets with values less than 0.001 percent for six of them.

How did individuals managed to perform so well? An analysis indicates that financial experience seems to have no correlation with game performance. The authors conjecture that feedback which leads to learning and adaption is the most important factor of success. A subject wrote: *'Admittedly, when first viewing the two data sets in the practice mode, it is impossible to tell which one is real, and which one is random, however, there is a pattern that quickly emerges and then the game becomes simple and the human eye can easily pick out the real array (often in under 1 second of time).'*

It is well known that computers struggle with many image-recognition and classification tasks that are trivial for humans. This is a possible explanation for the individuals performance. We argue that although the test shows that people are successful in identifying the difference between random and real time series no conclusion can be made about the ability of individuals in forecasting prices.

3.1.5 Risk and Return

Risk and return are the starting point for any investment decision in Modern Portfolio Theory. A major assumption is that risk and return are positively related to each other - it is not wise to consider them independently and more risk goes in line with more return and vice versa. Risk is in investment often measured by the variance, i.e. the weighted average of squared deviations from the mean. Since variance is expressed in square units and difficult to interpret one prefers the square root as measure of risk - the standard deviation or **volatility**. Since the volatility measure stick punishes also the positive deviations from the mean and fails to measure the heavy impact events (tails), today one uses often asymmetric risk measures such as Value-at-Risk. But since for liquid asset classes the return on longer investment horizons can be considered to be normally distributed, VaR is given by a multiple of the variance. We therefore continue to work with the variance. Assets from different asset classes show different volatility. To make return comparable, one adjusts them for the risk. The Sharpe ratio is such a measure defined as excess return, i.e. return over risk free rate, normalized by the annualized volatility of excess return. If this ratio is close to zero for an asset there is no long term return premium, i.e. it does not pay to keep an unhedged position. The real risk free rate represents the time value of money or the opportunity cost of trading current for future consumption.

Calculating the risk-return relationship for different asset classes shows that the trade-off **changes** over time. The changes can be classified into three different economic events. First, economies show mean-reverting behavior. The economy **fluctuates** around the mean-reversion. In a recession interest rate are cut to stimulate business activity and volatility increases due to higher default risk and increasing uncertainty about future firm profitability. Second, **structural shifts** impact the risk-return trade-off. Japan faced such a shift in the 90's of last century and the Euro Crisis is likely to reprice many European assets to a new equilibrium value. Third, **event risks** such a financial crisis, new regulations effect the risk-return trade off.

Given the risk free rate and risk premia, we can use them to forecast expected asset returns. How should this be done? We could extrapolate the historical values. This is not a good idea since it postulates that the past repeats itself: Falling or increasing value will continue to do so. Hence, we fail to account for cyclicality by extrapolating past trends. A more suitable approach is to construct the returns bottom-up by comprising stationary and variable components, see Figure 3.6 for an example.

The approach of Modern Portfolio Theory (MPT) follows the principle to obtain the highest possible return given a level of risk. The elements which matter in MPT are the expected return, the volatility of the individual assets and the co-movement of asset prices (covariance). Consider two assets, a bond and a stock. The expected return is the

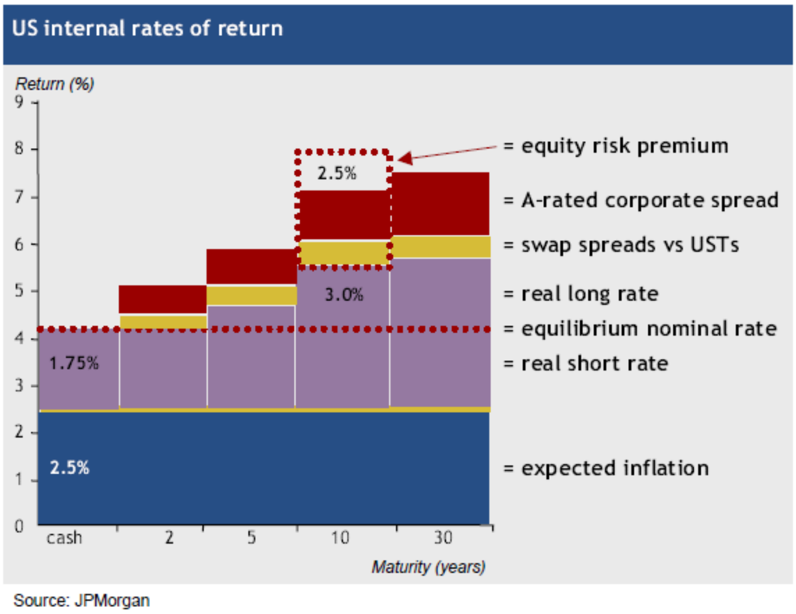


Figure 3.6: Forecasting long-term return using a bottom-up approach.*Source: JP Morgan, 2004*

weighted average

$$E(R) = 0.6 \times 5\% + 0.4 \times 4\% = 4.6\%$$

where we assume that 60 percent of the wealth is invested in the stock, 40 percent in the bonds and the expected returns are 5 and 4 percent, respectively. Expected return is additive, i.e. co-movements do not matter. The risk dimension variance or volatility is not additive, i.e. covariance matters in the risk of the assets:

$$\mathrm{cov}(R_{\text{stock}}, R_{\text{bond}}) = E[(R_{\text{stock}} - \mu_{\text{stock}})(R_{\text{bond}} - \mu_{\text{bond}})]$$

with μ the expected returns. If one asset is risk free, covariance is zero. Since covariance is not bounded this makes it difficult to interpret. Normalizing covariation by the product of the two assets volatilities correlation follows which takes value between -1 (perfect negative correlated) and $+1$ (perfect positive correlated). Positive correlation means that if stock price moves in one direction then also the bond price moves without delay in the same direction and vice versa. The standard rule for the variance of a sum of random variables (portfolio) implies that this variance is equal to the weighted individual variance's sum plus their correlation. Hence, a negative correlation value reduces the sum of the individual risk components for portfolio risk keeping the return level unchanged. This is the value added of diversification. Unfortunately, most asset classes show positive correlation and correlation is not stable over time, see Figure 3.7. The portfolio choice 60/40 was arbitrary. If we draw the expected portfolio returns and the portfolio standard deviation, for different portfolios different points follow. The maximum risky portfolio 100/0 and the lowest risky one 0/100 are shown. The straight line represents all possible portfolio choices if there is perfect positive correlation. If we fix any portfolio on this line and consider portfolios which give the same return with higher risks or for the same risk provide lower return, we call this portfolio a **dominate** one and the portfolio which dominates such a choice is called an **efficient** portfolio. The set of these portfolios is the efficient set or efficient frontier. If we vary the correlation parameter, the efficient frontier varies. The frontier becomes maximally bowed if perfect negative correlation holds: The lower correlation is the higher are the gains from diversification. So far, all asset were assumed to be risky. We add a risk less asset to the picture, i.e. an asset with zero volatility by definition. Hence for 100 percent in the risk free asset the efficient frontier starts on the y-axes. The Mutual Fund Theorem (see below) states that every efficient portfolio can be written as a linear combination of two efficient portfolios. Therefore, we can obtain an profile for portfolios on the straight line (Capital market line) which is tangent to the efficient set of risky assets only. The tangent point represents the highest attainable portfolio of risky assets - the so-called **market portfolio**. At the tangent point the investor holds a combination of cash (risk less asset) and the market portfolio. Regardless of where an investor is positioned on the capital market line he holds the market portfolio. The proportion of each asset in the portfolio are the same for all investors. Therefore the efficient portfolio is a market capitalization weighted basket of securities. Hence in theory all investors hold the market portfolio. The assumptions are that all investors perceive risk equivalently, all investors hold a combination of cash and the market portfolio, a market-cap weighted basket of securities is the efficient portfolio,

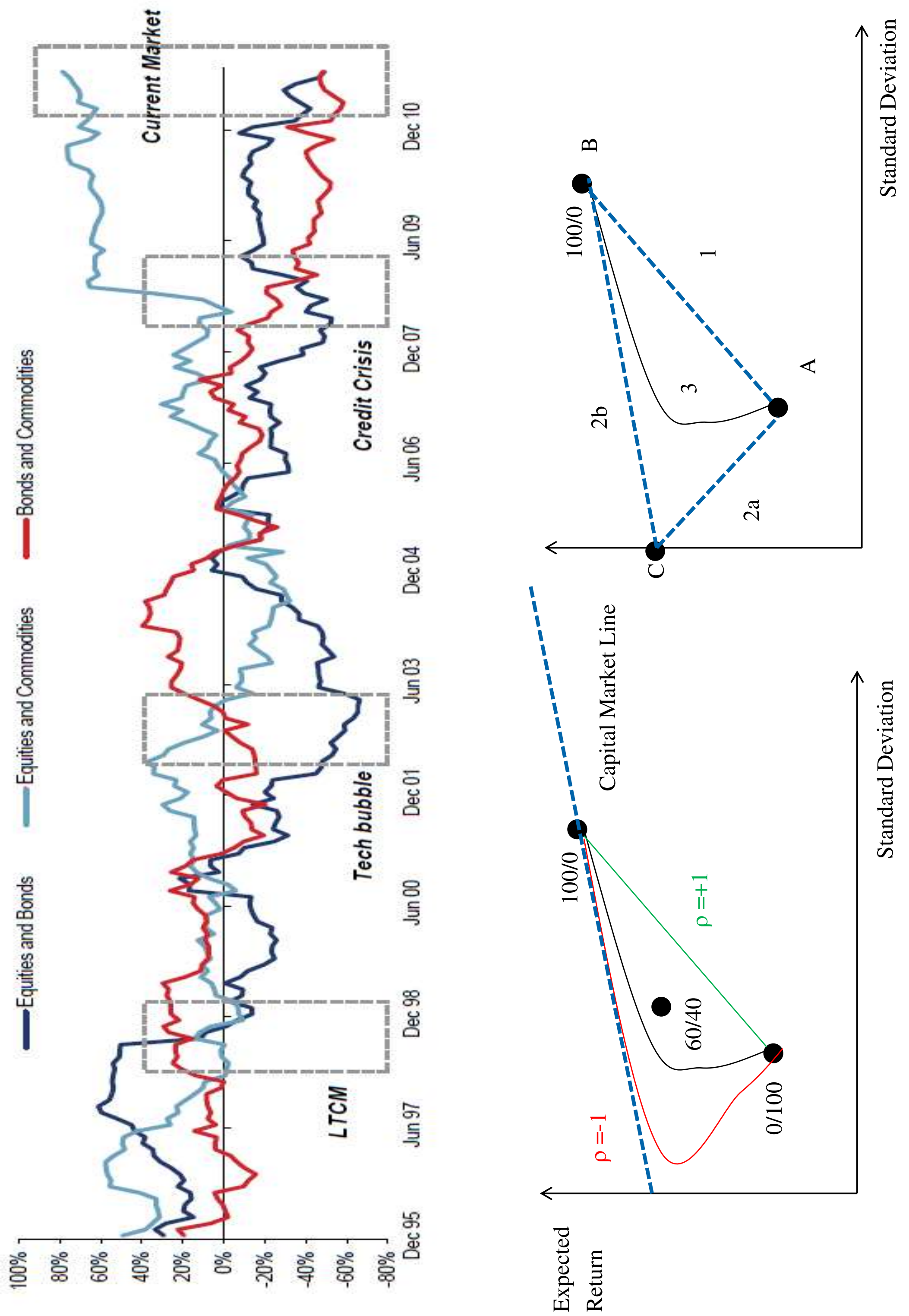


Figure 3.7: Upper Panel: Time varying correlation.*Source: Goldman Sachs, 2012.* Lower Panels: Efficient frontiers (sets) as a function of asset correlation.

there in no compensation for non-market allocations and there is not point for active management, i.e. if the market portfolio is efficient, there is no role for active investing.

We consider the two asset example im more details. Let $\phi = (\phi_1, 1 - \phi_1)$ be the corresponding normalized portfolio of an investment in **two risky assets**. Where can the feasible portfolios be positioned in the risk-return space with coordinates (σ_ϕ, μ_ϕ) of a portfolio ϕ with

$$\begin{aligned} \mu_\phi &= E[R^\phi] = \langle \phi, \mu \rangle = \phi_1 \mu_1 + \phi_2 \mu_2 \\ \sigma_\phi^2 &= \mathrm{var}(R^\phi) = \phi_1^2 \mathrm{var}(R_1) + \phi_2^2 \mathrm{var}(R_2) + 2\phi_1\phi_2 \mathrm{cov}(R_1, R_2) \ ? \end{aligned} \tag{3.3}$$

To find the relationship between risk and return we solve the variance equation w.r.t. to the strategy weight ϕ_1. This gives two possible solutions $\phi_1^\pm$

$$\phi_1^\pm = \frac{B \pm \sqrt{D}}{A + B} , \tag{3.4}$$

where

$$A = \sigma_1^2 - \rho\sigma_1\sigma_2 \qquad B = \sigma_2^2 - \rho\sigma_1\sigma_2 \ , \quad D = B^2 - (\sigma_2^2 - \sigma_\phi^2)(A + B) \tag{3.5}$$

Inserting the solutions $\phi_1^\pm$ in the return expression (3.3) for μ_ϕ we obtain $\mu_\phi^\pm$:

$$\mu_\phi^\pm = \mu_2 + \phi_1^\pm(\mu_1 - \mu_2) \ . \tag{3.6}$$

Thus, μ_ϕ is as a function of σ_ϕ an hyperbola where for any possible level of standard deviation σ_ϕ two portfolio expected returns $\mu^\pm$ are generally obtained. Without loss of generality we assume $\mu_1 < \mu_2$ and $\sigma_1 < \sigma_2$.

In the case $D = 0$, i.e. $D = 0 \iff \sigma_\phi = \sqrt{\frac{\sigma_2^2 - B^2}{A+B}}$, a single portfolio $\phi = \phi^+ = \phi^-$ follows. This implies a single expected return solution

$$\mu^\phi = \frac{\mu_2 A + \mu_1 B}{A + B} \ . \tag{3.7}$$

This portfolio is the **global minimum variance portfolio**, i.e. the portfolio attaining the minimal variance risk in the set of all feasible portfolios. Since the discriminant D is monotonically increasing in σ_ϕ, because $A + B > 0$. Hence the minimum $D = 0$ is attained at the minimal allowed standard deviation σ_ϕ. The corresponding portfolio ϕ is therefore the global minimum variance one.

The portfolio opportunity set is a hyperbola in $(\sigma_\phi, \mu_\phi)-$coordinates. If the risky assets are **perfectly** (positively or negatively) correlated, the relationship between the portfolio return standard deviation and if the expected portfolio return is **linear**, see Figure 3.7 (lower panel right). The point A represents asset 1; B represents asset 2. Asset 1 has lower expected return and variance than asset 2. Consider the case $\rho = 1$. Then *all* efficient portfolios lie on a straight line 1 in the mean-standard deviation space.

Any risk and return combination between the ones of assets 1 and 2 is obtained as a portfolio $\overline{\phi} = (\overline{\phi}_1, 1 - \overline{\phi}_1)$ determined by

$$\mu^{\overline{\phi}} = \overline{\phi}_1\mu_1 + (1 - \overline{\phi}_1)\mu_2 \ ,$$

for any given target expected return $\mu_1 \leq \mu^{\overline{\phi}} \leq \mu_2$. The standard deviation $\sigma^{\overline{\phi}}$ then follows. Similarly, for $\rho = -1$ all portfolios expected returns and standard deviations will lie either on the straight line $2a$ or on the straight line 2b, respectively, depending whether $\mu^{\overline{\phi}}$ is larger/smaller than $(A\mu_2 + B\mu_1)/(A + B)$, i.e. the expected return of the minimum variance portfolio. In this case the minimum variance portfolio is a risk free asset, since it has zero standard deviation of the return. In the presence of a perfect negative correlation we can fully eliminate the portfolio risk while having long positions in both assets. In such a setting, asset 1 is a perfect hedging instrument for asset 2 (and vice versa). When comparing for any target expected return $\mu^{\overline{\phi}}$ the standard deviation implied for $\rho = -1$ and $\rho = 0$ one can observe the risk reduction deriving from the stronger hedging effects arising when $\rho = -1$. More generally, hyperbolas for lower $\rho-$values will tend to be moved to the left, in mean-standard deviation space, due to the stronger hedging effects obtained in this case.

The portfolio returns of portfolios of more than two risky assets will be **scattered** on a whole surface in mean-standard deviation space.

Definition 3.1.5. *Assume that investor's preferences on portfolio returns depend only on the expected return and the variance of the portfolio. The following terminology is used:*

- *If a portfolio offers a larger expected return than a second portfolio for the same risk, then the latter portfolio is strictly dominated by the first one.*
- *If a portfolio offers a smaller risk than a second portfolio for the same expected return, then the latter portfolio is strictly dominated by the first one.*
- *Portfolios that are not strictly dominated by some other one are called mean-variance efficient or simply efficient, i.e. we define that a portfolio ϕ^* as mean-variance efficient if there exist no portfolio ϕ such that*

$$E[R^{\phi}] \geq E[R^{\phi^*}] \ , \ var(R^{\phi}) < var(R^{\phi^*}) \ . \tag{3.8}$$

The Markowitz model allows us to **systematically** determine the set of all efficient portfolios.

One can argue at this point that the variance or volatility is not an adequate risk measure. This is indeed the case. There are more convenient risk measure sticks such as value-at-risk or even coherent or convex risk measures. We discuss them at the end of this chapter. Although there has been impressive research in the last decade concerning appropriate risk measures in investment besides the variance only value-at-risk seems to be used. The reason is that the dependence of the assets in markets under stress is becoming extreme - the benefits of diversification vanishes. It is important to have a view on different markets, to have a clear policy when to enter or exit the markets.

3.1.6 Indices

Indices are often used as benchmarks or as investment vehicles to achieve broad diversification within an asset class. Besides the member asset prices, there are four other main factors determining the index value. In order to calculate the index value the following factors have to be taken into consideration: member weighting, divisor, index return type and value fixing. In general the formula for index calculation looks like

$$I = \frac{\sum_{i=1}^{M} w_i S_i}{D}$$

where M is the number of assets in the index, S_i is the price of a tradable unit of asset i, e.g. the price of a stock, w_i is the weight assigned to the price of that asset and D is the divisor.
Weighting. Various methods are used for determining the weight of individual members in the index. Within the same category there can be subcategories and slight differences.

- Market capitalization weighting: The members are weighted proportional to the total market value of the asset issuer, i.e. Wi is dependent on the size of the company for equity. In the equity case this would correspond to the number of outstanding free-floating shares multiplied by the share price. Subgroups of this weighting would be if weights were capped at some level, or that no consideration was taken into free float. This is the most common form of weighting for public indices and the rule for indices such as S&P, FTSE, MSCI and SMI .

- Equal weighting 1 (Price Weighting): The weight assigned to different assets is the same. As a consequence the price of a tradable unit of the asset will have a determining effect on the weight of an asset in the index . Dow 30 and Nikkei 225 indices are calculated using the equal weighing scheme.

- Equal weighting 2 ('CHF Weighting'): The CHF weight assigned to each asset is the same, i.e. $S_i w_i$, is the same for each asset. This means that if CHF 500 is to be invested in a basket of 10 assets, the amount bought of each asset would be CHF 50.

- Share weighting: The members are weighted proportional to the total number of tradable units issued, i.e. Wi is dependent on the number of the shares outstanding for the equity asset class.

- Attribute weighting: The members are weighted according to their ranking score in the selection process. If our ranking is based on ethical and environmental criteria, and asset Y has a score of 75 and asset X 25, then weight ratio between asset Y and X will be Weight Y / Weight X = 3.

- Hybrid or Custom weighting: The weighting scheme can be a combination of the above alternatives or be something totally new, maybe based on the request of client.

Free-floating is the portion of total shares held for investment purposes. This is opposite to shares held for strategic purposes, i.e. for control. Some indices are quoted using different weighting schemes, e.g. MSCI. However, the main quoted value is using the market capitalization weighting method.

Remark:
The difference between the asset weighting scheme and the weight of an asset in the index is as follows. For a price weighted index $w_1 = w_2$ for asset 1 and asset 2. However if $S_1/S_2 = 3$ the weight of asset 1 in the index will be 3 times larger than the weight of asset 2.

Divisor
The divisor is a crucial part of the index calculation. At initiation it is used for normalizing the index value. The initial SMI divisor on June 1998 was chosen to a value which normalized the index to 1500. However, the main role of the divisor is to remove the unwanted effects of **corporate actions** and **index member change** on the index value. It ensures continuity in the index value in the sense that the change in the index should only stem from the investor sentiment and not originate from "synthetic" changes. Corporate actions, which need to be accounted for by changing the divisor value, are dependent on the weighting scheme used for the index. Consider a stock split for an index with ...

- ... market capitalization weighting: The price of stock will be reduced and the number of free-floating shares increases. These two effects will be offsetting and no change has to be made to the divisor.

- ... equal weighting (Price Weighting): The stock price reduction will have an effect, but the number of free-floating share has no impact on such a weighting. Therefore, the divisor has to be changed to a lower value in order to avoid a discontinuity in the index value.

It is very important in practice to have a good understanding of the influence of common corporate actions such as splits, dividends, spin off, merger & acquisition, rights offering, bankruptcy, etc on the index value so that the index value continuity can be ensured.

Return Type

How the dividends are handled in the index calculation determines the return type of the index. There are three versions of how dividends can be incorporated into the index value calculations:

- Price return index: No consideration is taken to the dividend amount paid out by the assets. The day-to-day change in the index value reflects the change in the asset prices.
- Total return index: The full amount for the dividend payments is reflected in the index value. This is done by adding the dividend amount on the ex-dividend date to the asset price. Thus, the index value acts as if all the dividend payments were reinvested in the index.
- Total return index after tax: The dividend amount used in the index calculation is the after tax amount, i.e. the net cash amount.

Value Fixing

Another set of rules that characterize an index calculation, is the data values and the frequency, which they are used. An index value is usually calculated in real time or once a day. Exceptions are illiquid indices for real estate asset class as an example.

The values that are needed for index value calculation can be quoted in various versions. The most important value is the asset price. It has to defined weather the value uses is mid prices, bid or ask prices, last trade prices or any other price value provided.

In addition, if the index constituents have a wide geographical span, there are other issues that need to be taken into consideration. Some of the rules that need to defined are: index value quotation currency, source of currency rates, index opening and closing hours, and assets registered on multiple exchanges.

For most major indices the quotation is real time and the currency rate used is also real time. The opening hour for the constructed index starts with the opening of the exchange of any index member, and the closing occurs when no index member exchange is open. Having a global index, with constituents from Japan to USA, would mean that the index would be "open" most hours of the day.

3.2 Beta: Delta One and ETF

Delta One represents investments where the sensitivity w.r.t. to the underlying assets is one - i.e. linear payoff profiles. There are many investment motivations and many products in Delta One business.

Investment motivations are reductions of productions costs, optimization of investment return and risk, gaining access to specific markets, diversification, taxation issues, managing liquidity and leverage constraints, term financing, managing cash flows,

cash optimization and leverage, precise hedging or investing in strategic themes.
There is a wide range of products serving the investor's purpose.

- **Stocks**. Direct investment in stocks has the advantages of customization and flexibility. It requires an infrastructure and one should anticipate a low turnover. Infrastructure is expansive and borrowing costs for the stocks can add significant variable costs.

- **Single or index stock swaps.** An equity swap exchanges a floating interest rate (LIBOR) and the performance of either a stock or a stock market index. These products are used to gain leveraged access to (global) markets and to access markets which are otherwise difficult to trade since thy require only initial and variation margins. They are used also in markets where stand liquid products may not exist. These products provide a taylor-made risk and return profile at low production costs since there is no management fee and administration as well maintenance costs are low. Swaps are best suited for large portfolios, long term trades and low turnovers. Other types of swaps are sector swaps and term swaps. Swaps possess low fixed costs (bid-offer, commissions) but face some variable costs from counterparty -, funding risk.

- **Certificates ([Retail] Structured Products).** These products allow for access to countries or products that may be otherwise inaccessible. They can be used by investors who have restrictions to trade derivatives. Certificates can be structured as short, long or a combined exposure. For institutional investors they can be completely be taylormade - payoff, underlying value and wrapping. The certificate composition and intraday prices are known at any time. Certificates are ment for longer term investments and low turnover. Besides bid-ask spreads and commissions, management fees applies. The most important variables costs are from issuer risk.

- **Futures** are best suited for high turnover and shorter duration. The fees are low (bid-ask, commissions) but they face variable costs from rolling -, dividend -, tracking error and funding risk.

- **ETF** are liquid, easy to use, are well suited for high frequent trading and if a more granular exposure is wished than for example futures allow. The fix costs consists of commissions, bid-ask spread and management fees. The all-costs are for retail type ETF between 20 and 50 bps p.a. and between 5 and 15 bps p.a. for institutional investors. ETF face tracking error risk and depending on the structure of the ETF dividend treatment. ETFs exist since the early 1990s as a cost- and tax-efficient alternative to mutual funds. The structuring of these funds initially shared common characteristics with that of mutual funds.

We consider ETF in more details. Mutual funds use ETF on a strategic level, Hedge Funds also on a tactical one. Initially ETFs shared common characteristics with that of mutual funds, i.e. the underlying index exposure that the ETF replicated was gained by

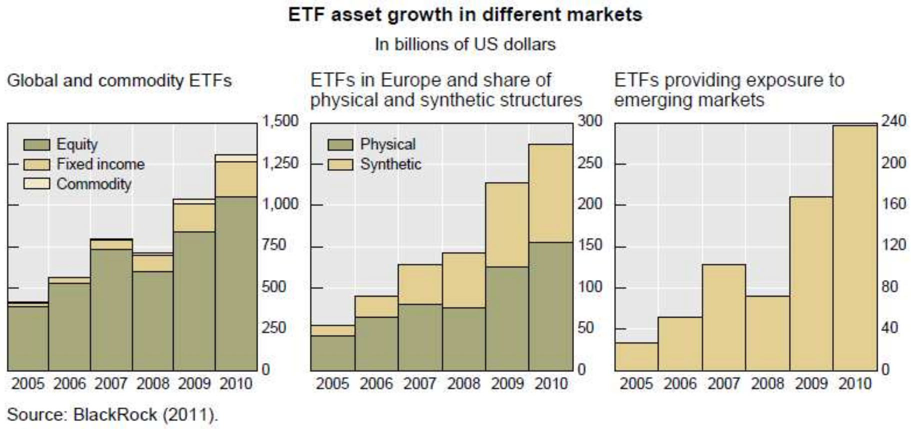


Figure 3.8: ETF asset growth in different markets. *Source: Ramaswamy (2011), BIS, and BlackRock (2011).*

buying the physical stocks or securities in the index. In recent years, investors desire to seek higher returns by taking exposure to less liquid emerging market equities and other assets through ETFs that guarantee market liquidity has demanded more innovative product structuring from financial intermediaries. ETFs have moved away from being a plain vanilla cost- and tax-efficient alternative to mutual funds to being a much more complex and diverse array of products and replication schemes. Table **??** some figures of the ETF industry.

Number ETFs	2’500
Sponsors	130
6 Main Sponsors	iShares, State Street Global Advisors, Vanguard, Lyxor Asset Management, db x-trackers and Power Shares control over 80% ETF market shares
Assets under management	1.3 Billion USD
Europe	80% of ETF assets in Europe are held by institutional investors
US	50% of ETF assets in Europe are held by institutional investors

Table 3.2: Some figures of the ETF industry. *Source: Blackrock, 2011*

Figure 3.8 shows the ETF asset growth.

How are ETFs constructed or what defines the ETF value chain?

- **Manufacturing**. ETF are constructed by index providers, exchanges, regulators, investment bank which apply the swap structuring, index fund managers.
- **Distribution**. Manufacturing is linked to the sales channel. Sales can be ETF-specializes sales, investment banking sales, asset management sales or third party distributors.
- **Clients** are mutual funds, hedge funds, institutions or private bank clients.

The construction of ETF can be broadly done in two ways: Direct replication or swap-backed construction. In the direct replication one buys all index components (full replication) or just an optimized sample. This is a transparent approach with low counterparty risk due to securities lending. But if the markets are difficult to access, significant tracking error can realize. In the swap-based approach one invests indirectly in a basket by achieving the index performance via a total return swap, see Figure 3.9. This approach minimizes the tracking error and one can access more underlyings. The drawbacks are counterparty risk and documentation (ISDA).

Which method will dominate in the future depends on two factors: Clients need and regulation. Since regulation impacts the costs, the clients need are not independent on how regulation will be defined. The swap approach will be stronger regulated where regulation will ask for more transparency. That is transparent on the collateral posted by the swap counter parties and on the securities lending process.

Figure 3.10 shows how private banks can use ETFs. The bank seeks for ETFs which can be used to construct a portfolio based on the building blocks Regional, Single Country, Sectors and Themes. The bank uses such an ETF portfolio complementary to the active stock-picking strategy. Compared to the active strategy this passive component has very low costs.

The cost components of an ETF are the spread, the broker charges and trading commission and the Total Expense Ratio (TER). The spread is driven by several components. Market liquidity has an impact, taxes such as stamp duty on underlying components of physically replicated ETFs, execution costs such as brokerage fees on underlying components in some markets (e.g. 30bps sell tax in Korea), creation or redemption cost and tracking error due to restricted stocks or closed markets. The TER is the sum of the management fee and the expenses. The expenses depend on the complexity of the underlying portfolio and the replication costs. For plain vanilla structure the TER can be 10bps but rise to 100bps for complex structures. TER is deducted daily from NAV.

Synthetic ETFs allow replication of the index using derivatives as opposed to owning the physical assets. One motivation for using synthetic structures to replicate the index could be to reduce costs. Physical replication can be an expensive method for tracking broad market indices such as emerging market equity or fixed income indices, or other less liquid market indices. Including only a subset of the underlying index securities for physical replication can lead to significant deviation in returns between the ETF and the index in volatile market conditions (tracking error). One popular synthetic

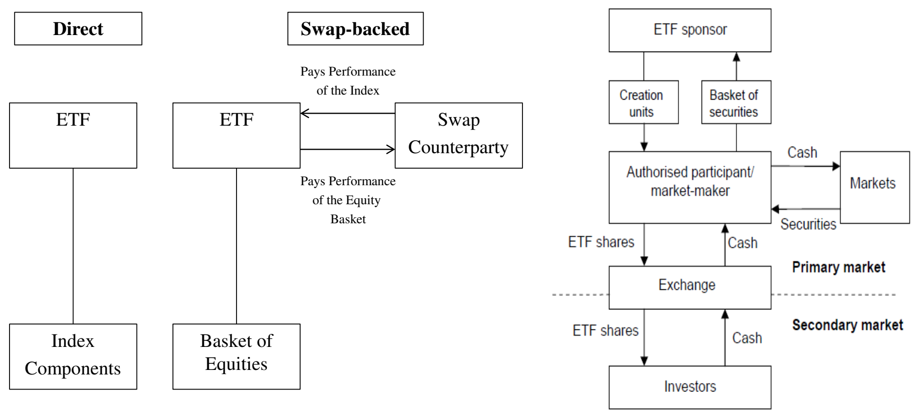


Figure 3.9: Left Panel: ETF construction. Right Panel: Operational structure of ETFs. Market makers purchase the basket of securities in the markets that replicate the ETF index and deliver them to the ETF sponsor. In exchange each market-maker receives ETF creation units (50,000 or multiples thereof). The transaction between the market-maker and ETF sponsor takes places in the primary market. Investors who buy and sell the ETF then trade in the secondary market through brokers on exchanges. The market value of the basket of securities held by the ETF sponsor forms the basis for determining the NAV of the ETF held by investors.*Source: Ramaswamy (2011), BIS*

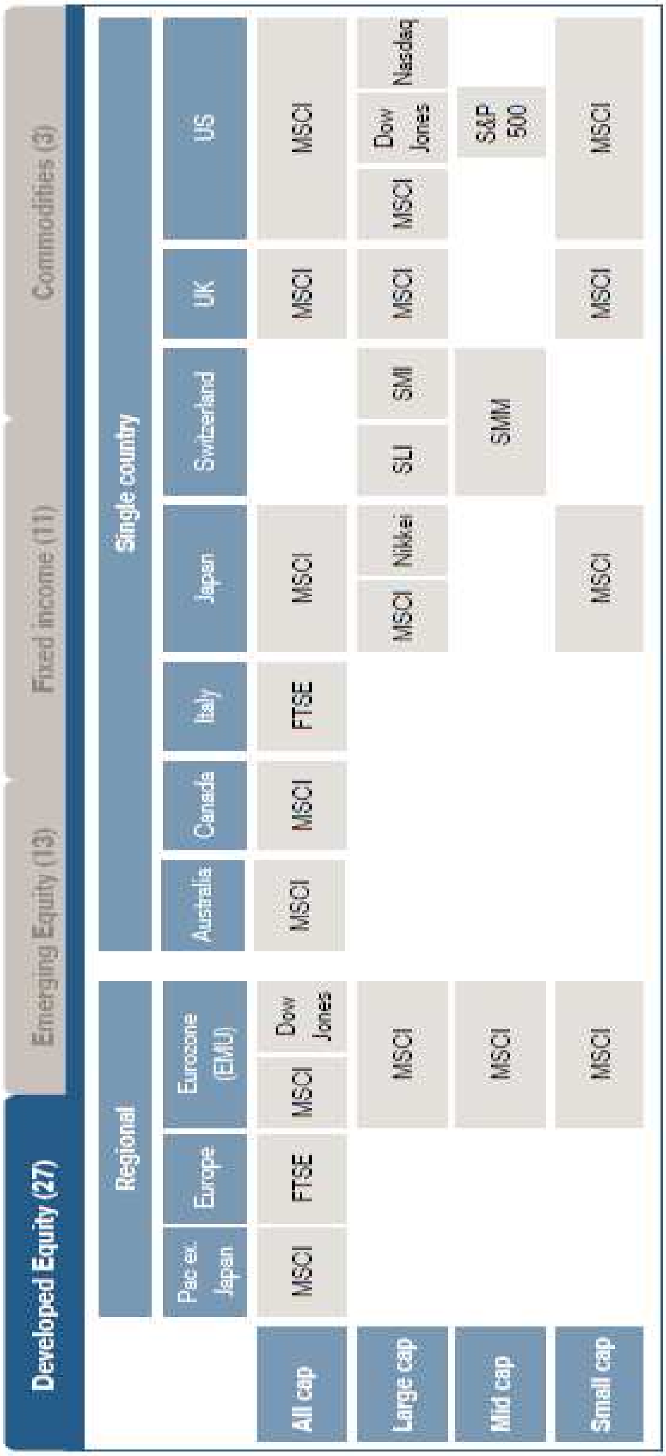


Figure 3.10: ETF portfolio via building blocks. MSCI Global Equity Indices are widely tracked global equity benchmarks and serve as the basis for over 500 exchanged traded funds* throughout the world. FTSE calculates over 120,000 end of day and real-time indices covering more than 80 countries and all major asset classes. Nikkei 225 index is a stock market index for the Tokyo Stock Exchange. S& P 500 is a free-float capitalization-weighted index based on the common stock prices of 500 American companies. The NASDAQ-100 Index includes 100 of the largest domestic and international non-financial securities listed on The Nasdaq Stock Market based on market capitalization. SLI Swiss Leader Index consists of the SMI stocks and the 10 largest SMIM stocks. *Source: Credit Suisse, 2010.*

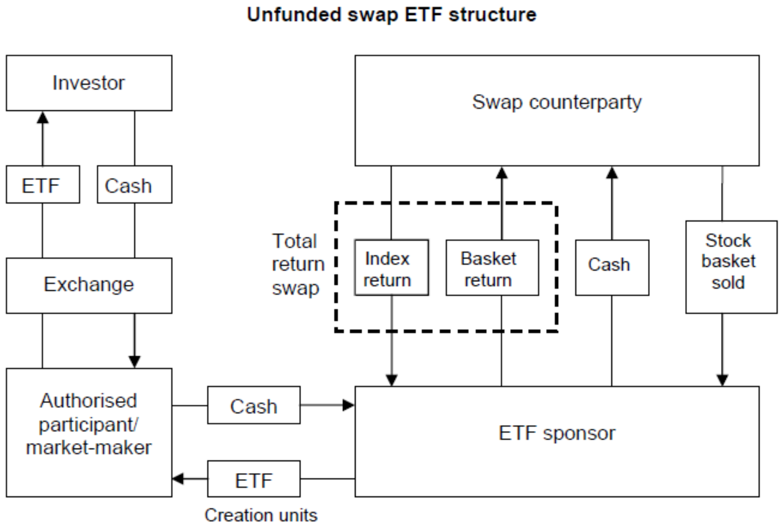


Figure 3.11: Unfunded swap ETF structure. *Source: Ramaswamy (2011), BIS*

structure involves the use of total return swaps which the ETF sponsors refer to as the unfunded swap structure, see Figure 3.11. Ramaswamy (2011) states*:'Under the synthetic replication scheme, the authorized participant receives the creation units from the ETF sponsor against cash rather than a basket of the index securities as in the physical replication scheme. The ETF sponsor separately enters into a total return swap with a financial intermediary, often its parent bank, to receive the total return of the ETF index for a given nominal exposure. This constitutes the first leg of the swap. Cash is then transferred to the swap counterparty equal to the notional exposure. In return, the swap counterparty transfers a basket of collateral assets to the ETF sponsor. The assets in the collateral basket could be completely different from those in the benchmark index that the ETF tries to replicate. The total return on this collateral basket is then transferred to the swap counterparty, which constitutes the second leg of the total return swap.'*

The nature of the swap transaction exploits synergies between banks' collateral management practices and the funding of their warehoused securities. Synthetic replication schemes transfer the risk of any deviation in the ETF's return from its benchmark to the swap provider, which is effected by entering into a derivatives contract to receive the total return of the benchmark. This protects investors from the tracking error risk which physical replication schemes would otherwise expose them to. However, there is a trade-off: the lower tracking error risk comes at the cost of increased counterparty risk to the swap provider. The increased popularity of ETF products among investors has led to greater competition between ETF sponsors, forcing them to seek alternative replication techniques to optimize their fee structures. Ramaswamy (2011) states:*'One outcome of this fee structure review has been to explore the scope for possible synergies that might exist between the investment banking activities of the parent bank and its asset management subsidiary or the unit within the parent bank that acts as the ETF sponsor. These synergies arise from the market-making activities of investment banking, which usually require maintaining a large inventory of stocks and bonds that has to be funded. When these stocks and bonds are less liquid, they will have to be funded either in the unsecured markets or in repo markets with deep haircuts. By transferring these stocks and bonds as collateral assets to the ETF provider sponsored by the parent bank, the investment banking activities may benefit from reduced warehousing costs for these assets. Part of this cost savings may then be passed on to the ETF investors through a lower total expense ratio for the fund holdings. The cost savings accruing to the investment banking activities can be directly linked to the quality of the collateral assets transferred to the ETF sponsor. For example, there could be incentives to post illiquid securities as collateral assets. Typically, such securities will have to be funded by the investment bank at unsecured borrowing rates. By posting them as collateral assets to the ETF sponsor in a swap transaction, the investment bank division can effectively fund these assets at zero cost for its market-making activities. In addition, the bank providing the total return swap through the unfunded swap ETF structure may benefit from a reduction in regulatory capital charges. This would be the case if lower credit quality and less liquid assets are included in the collateral basket sold to the ETF sponsor compared with those acquired for replicating the ETF index.'* In Ireland, for example, equities posted as collateral are subject to a 20 percent haircut, whereas in Luxembourg it is up to the fund custodian and the fund management company to negotiate the haircut. As a consequence, UCITS-compliant ETFs that are synthetically replicated tend to be registered in Luxembourg to reduce haircuts on collateral assets posted.

An alternative replication scheme used by ETF sponsors is to employ the so-called funded swap structure. Under this, the ETF sponsor transfers cash to the swap counterparty, who then provides the total return of the ETF index replicated. This transaction is collateralized, with the swap counterparty posting the eligible collateral into a ring-fenced custodian account to which the ETF sponsor has legal claims. This structure is less commonly used by sponsors for synthetic replication of ETF indices

3.3 Alpha: Hedge Funds

What is a HF? HF allow for collective investments. In this sense they follow the same purpose than ordinary investment funds. But there are differences.

- Many HF are offshore domiciliated, i.e. on some islands or countries which offer them tax advantages and/or which have low regulation standards. But regulation of HF is changing. The EU agreed 2010 to apply stronger regulatory rules for HF. The G20 agreed in 2009 to regulate the 100 largest HF. The US SEC an the British FSA have then the right for insight in the HF balance sheets.

Comparison of the two ETF synthetic replication schemes

Attribute	Unfunded swap structure	Funded swap structure
Type of transaction	Total return swap	Credit- or equity-linked note
Counterparty risk	Mitigated through ownership of collateral assets effected through true sale	Mitigated through the pledge of collateral assets through triparty agreement
Overcollateralisation	Typically small, if any	Typically up to 120% of the ETF's net asset value
Haircuts on collateral assets	Usually none, but under UCITS some jurisdictions can impose haircuts for certain assets	Usually none, but under UCITS some jurisdictions can impose haircuts for certain assets
Composition of collateral assets	Can change daily	Can change daily
Balance sheet implication for swap counterparty	True sale of collateral assets can alter risk-weighted capital charges	Pledge of collateral assets unlikely to alter risk-weighted capital charges

Figure 3.12: Comparison of the two ETF synthetic replication schemes. *Source: Ramaswamy (2011), BIS*

- Many HF cannot be offered to the public, i.e. private placement to qualified investors often defines the client and distribution channel.
- An important selling argument for HF is that their investment not or only weakly correlate with traditional markets. This argument needs more explanation. First, HF often invest in traditional markets too, i.e. the above argument means than that their investment strategy in the traditional markets is only weakly correlated to traditional strategies in these markets. To what extend the HF argument holds true is changing over time. In the year 2000 and the following years, correlation between MSCI World, a broadly world-wide diversified stock index, and the broad DJ CS Hedge Fund Index or correlation between DJ-UBS Commodity Index and DJ CS Hedge Fund Index changed on a 2y rolling basis. The correlation was at 0.54 respectively at 0.16 between the commodity and the HF index in the years 2000-2007. 2007, correlations jumped to 0.8 and 0.81. A significant part of the HF managers started in 2007 to invest traditionally in stocks and commodities.
- Fee structure. HF often charge a fee of 1 percent per year of the NAV plus a performance fee. These performance fees can be as high as 20 percent of net income in a period.
- Investment strategies. HF use short selling and leverage strategies. In the latter one, they hope that funding costs of capital are lower than the return on the

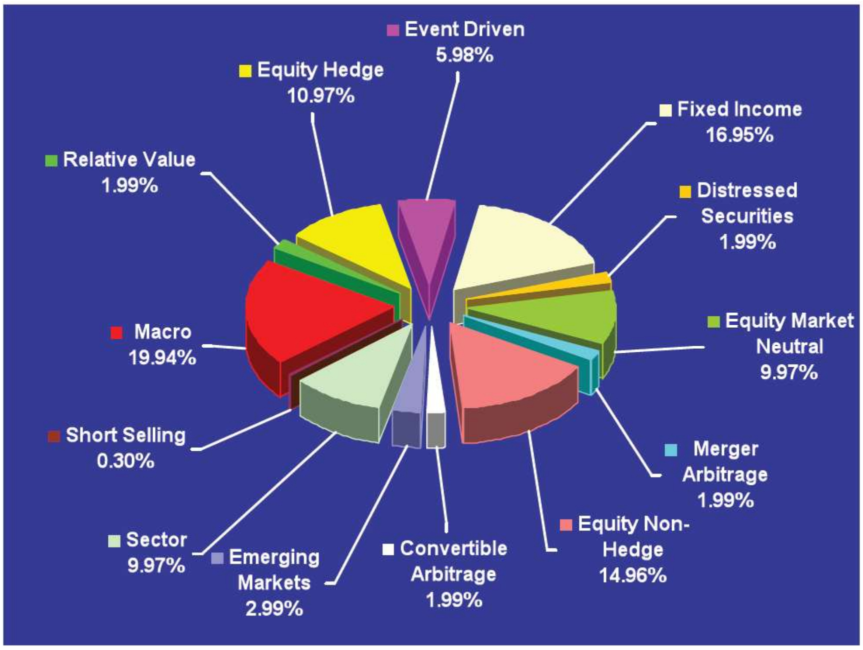


Figure 3.13: Fractions of different HF strategies. *Source: L. Seco (2010).*

leveraged capital. Some HF strategies are described next, see Figure 3.13:

- **Long-short stock strategies**, i.e. one is long the stock which one consider to be undervalued and short the overvalued one.
- **Relative Value or Arbitrage Strategies**. These strategies use mispricings between securities. For example, if a stock trades at different prices at two exchanges.
- **Event Strategies focus** on particular events which can effect specific firms, sectors or whole markets. This can be spin-offs or joint-ventures for firms or liquidity crunches for whole markets.
- **Global Macro Strategies** try to identify global economic trends and to replicate them using financial products. An example is the HF Quantum of G. Soros. This HF noted 1992 the overvaluation of the British Pound. Using hugh capital amounts the HF forced the Bank of England to stop to maintain the Pound - the Pound strongly depreciated against other leading currencies, the HF made large gains and UK was forced to leave ECU; the ancestor of the European Monetary Union.

We consider a **convertible arbitrage strategy** where we follow Seco (2010). A **convertible bond** has the following properties:

- It trades at USD 80 at the moment, i.e. below par. The coupon is USD 4 and the

Title	0	T_1	T_2	T_3	T_4	T_5	T_5
Bond	80	80	73	70	91	88	85
Stock	-70	-70	-60	-60	-80	-80	-80
T-Bill	+70	+ 70	+ 72.8	+ 72.8	+ 72.8	+ 72.8	+ 72.8
Coupon	−	+ 4	+ 4	+ 4	+ 4	+ 4	+ 4
Fee	−	+ 3.5	+ 3.5	+ 3.5	+ 3.5	+ 3.5	+ 3.5
Value	80	83.3	86.3	83.3	84.3	81.3	78.3

Table 3.3: Return of the convertible arbitrage strategy. The subscript x in T_x numerate the scenarios.

bond holder can convert the bond in 10 stocks before the bond matures. The price of the stock is 7.

- The bond holder sells the bond to the HF.
- The HF borrows 10 stocks for the price of USD 70 at a fee of 1 percent, sell the stocks for USD 70 and invests the amount in US treasury bills (T-Bills) with 4 percent return. The costs of selling the T-Bills and buying back the stocks is assumed to be 1 percent. This defines the convertible arbitrage strategy.

Table 3.3 shows the development of the P& L for different stock and bond price changes starting today at $t = 0$ up to one year at T. The table shows that in only one scenario a loss follows. This also shows that the strategy is **not** an arbitrage strategy in the strict academic sense. The strategy has several earning components. The bond yield, short-term rates, stock moves and the bond itself. The number of risk factors is considerable. Besides interest and credit risk, stock volatility, liquidity risk is the correlation between the bond and the stock market a relevant, difficult to estimate instable factor. The HF index for convertible arbitrage increased almost linearly between 1996 and 2005. During this period interest rates largely failed and after the dot.com bubble and 9/11 stock started to rise for a long period. This stock increase led after 2005 to falling volatilities. Since not only realized but also implicit volatilities became lower, the call option price which are based on the implicit volatilities became cheaper. But the right of conversion in a convertible bond is equivalent to an exposure in a call option. This led to a flattening of the index after 2005. In August 2008 of the financial crisis the index lost more than 40 percent of its value. The events in this period - the repricing of credit risk, the drop of stock markets, the squeeze in liquidity - correspond to a tail-event from a risk view. In such an event a structural change follows. Risks and their dependencies are newly calibrated. This shows that HF may provide better return properties for many scenarios than traditional fund investments but that the often more complicated strategies face higher losses is a structural change hits the HF.

While some consider HF to be a particular active investment strategy, others consider HF to be a own **asset class**. Since HF generally invest in traditional asset classes and markets and therein value added is expected to come from the skilful re-allocation of the portfolio among investment existing opportunities it is probably best to consider them as a complementary strategy to traditional strategies and not as an own asset class, see Figure 3.14.

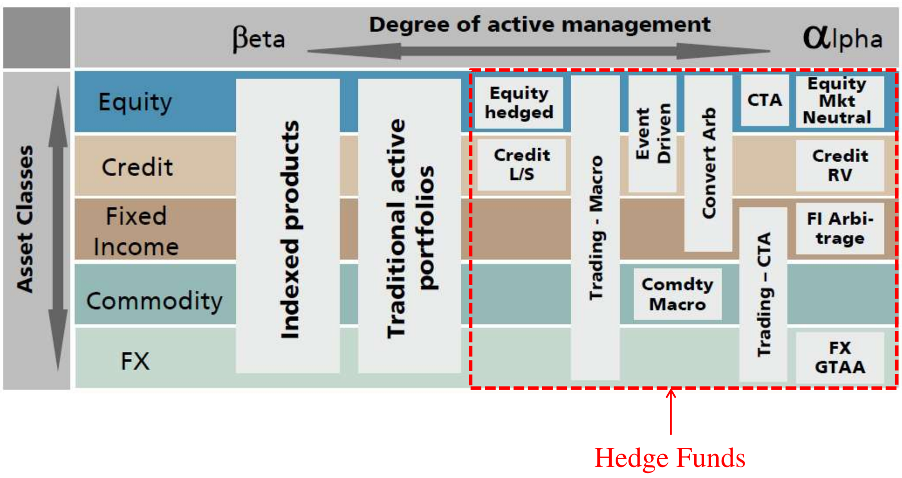


Figure 3.14: Asset classes and investment strategies. Credit L/S means credit long-short, CTA means Commodity Trading Advisors (CTA), where the fund trades in futures (or options) in commodity markets or in swaps, Credit RV means credit risk relative value strategies, FX GTAA means FX Global Tactical Asset Allocation. *Source: UBS Global Asset Management. 2011*

During the financial crisis the question was raised about the relevance of HF for systemic risk. Many authors such as the BIS in their annual report 2011 or B. Bernanke estimate that the risk is not large. Two reasons lead to their estimate. First, many HF are small in size. Second, many HF are not strongly connected to the banking sector.

The HF industry has a fund fortune of about USD 300 billion.[8] in 2000. This figure increased to USD 2'400 billion by the end of 2007. During the financial crisis the amount fall to 1'400 billion in 2008. 2010 the value was at 1'750 billion USD. The number of HF increased from 4'000 in 1999 to 9'400 in 2009. The average leverage factor varies around 1.5. This factor can increase for some strategies for a short period to values between 3

[8]The data are from the BIS annual report 2010, page 105 and the document of the City University, 2010, about HF.

and 4. More than 2/3 or all assets of HF are in the US, Europe follows with around 1/4.

The low regulation standard of HF compared to banks is often criticized. There are several causes for these debates. First, at the moment regulation of HF is weak compared to the banking sector. Second, HF often operate in a non-transparent way. This naturally raises suspicion. Third, given the number of HF and some prominent realized cases, it would be not natural if there were no black sheeps in the HF industry. If we restrict ourself to scientific work one notes the increasing number of papers which show evidence that the HF industry manipulates the performance reports, see Agarwal et al (2011), Bunnermeier et al (2004), Ben-David et al. (2012). The last cited work finds evidence that HF manipulate stock prices at their critical reporting dates. The stocks show at the last reporting date a significant excess return, which at the day after the reporting turns in the opposite. The authors further show that the incentives for manipulation increase for HF with stronger ranking incentives relative to other HF. Such manipulations affect all other market participants which invest in the same stocks.

3.4 Rule Based Investment

What is a rule? A rule is a mathematical description which states which title of a defined asset universe is selected at a given date for investment, what the amount of investment is in each title, how coupons payments are calculated and when do they are payed, how different styles such as momentum, volatility control are defined. Besides these type of mechanistic rules there also exist mixtures where some discretionary decision are allowed beside the rule mechanism.

What is the rationale to use mechanistic rules in investment? Rules are free of emotions. In a world such as the financial world which is complex, unexpectedly changing managers are sometimes less driven by rational analysis but more by emotions. Often, such situations lead to underperforming investments. Hence rule based investment assures that systematic thinking drives the investment process. On the other hand rule based models can never capture the full complexity of the real investment world; they are in particular blind to foresee structural changes. The attempt to make models more and more realistic increases their complexity and makes parameter estimates more and more difficult. Hence, to search for the most realistic and hence complex model is a dead end in investment.

Given the pros and cons of rule based models and discretionary judgement one sometimes sees a combination of them - model provide information for discretionary decision making. This takes away a fear to be model-driven but one remains informed by the models. To successfully combine these two approaches requires a lot of experience from the decision makers - discipline to control emotions when markets crash is an example.

3.4.1 Momentum and Volatility Control

Momentum strategy means buying stocks which had a positive return over a predefined past period and selling in the other case. Hence, momentum is basically a strategy which is related to the 'velocity' of asset prices. Whether or not shares are bought depends on the definition of the **signalling** process. Clearly, people which believe in the EMH will not invest according to a momentum strategy since for them returns are not predictable. The difficulty is that an increase in asset prices, in and of itself, should not warrant further increase. Such increase, according to the efficient-market hypothesis, is warranted only by changes in demand and supply or new information (cf. fundamental analysis).

Defining the momentum strategy requires to fix a time period and the signals over this period. If the period length is set too long the momentum strategy is likely detect only longer term trends. This would make sense even from an EMH point of view. Taking shorter periods the momentum strategy is on one hand side able to catch up also with smaller trends but the strategy also becomes more volatile, i.e. trading becomes more frequent. A simple example of a time horizon and a signalling strategy is for the weight $w_i(t+j)$ of investment in asset S_i over the period $[t, t+j]$:

$$w_i = \begin{cases} 1, & R_j = \frac{S_(t+j) - S(t)}{S(j)} > \overline{r}; \\ 0, & R_j = \frac{S_(t+j) - S(t)}{S(j)} \leq \overline{r}. \end{cases}$$

Momentum strategies are not successful if the markets are moving side way - then the risk is that the strategy always changes too late or too early the investments.

A different rule is **volatility control**. The rationale of volatility control is to bound the volatility of an investment portfolio, say at 10 percent. If one asset price falls, an empirical regularity implies that volatility of this asset increases (negative leverage). Therefore, volatility of the portfolio increases. Consider a second asset in the portfolio with low correlation to the first asset with different risk sources. Assume that the volatility of this asset is not increasing in the same magnitude as the first one does. Then the volatility cap on the portfolio implies that at a given level investment switch from the more volatile to the less volatile one. This if often considered to be beneficial for the investor since a drop by say 20 percent of the first asset requires an increase of 40 percent to recover the status quo. Therefore, even if there exists capital protection this is typically only effective at maturity date. Before, the investment can value well below the guarantee level. Using a volatility control one tries to circumvent heavy losses during the investment period. Consider the variance of a portfolio:

$$\sigma^2_{\text{port}} = \sum_j w_j^2 \sigma_j^2 + 2 \sum_{i<j} w_j w_i \rho_{ij} \sigma_i \sigma_j \leq \overline{\sigma^2}$$

with $\overline{\sigma^2}$ the variance cap level. Each value of this cap level defines multi-dimensional ellipsoids in the $w_j \sigma_j$-coordinate space, see Figure **??**.

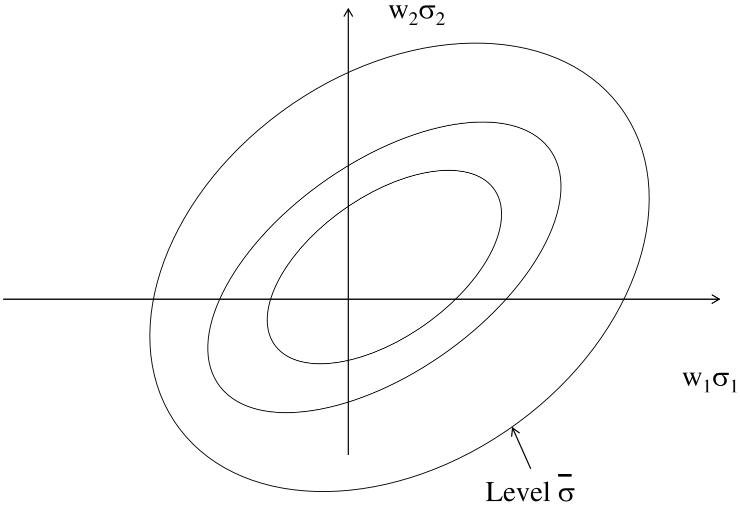


Figure 3.15: Portfolio variance level curves for two assets. On a given ellipse the portfolio variance assumes the same level.

Given the uncountable number of portfolio combinations which lead to the same target volatility, which one should one choose? At this point a second rule can help for example the momentum based rule. This leads to a preference ordering of the investment strategy w_j at each date. One can therefore choose that portfolio which respects the volatility constraint and which is optimal from the momentum strategy point of view.

3.4.2 CPPI

Option markets are basic to design structured products with capital protection and participation. But there are many underlying values without a sister option market such as many ETFs or investment funds. Is it possible to offer both protection and participation in these cases? Yes if we recall that a call or put option is a leveraged product. That is we only have to rebuild leverage. To understand the mechanics assume that an investor wishes capital protection and participation by investing in a fund. More specific, out of his CHF 100 investment

- CHF 75 are used to generate capital protection in 7 years and
- CHF 25 are free to define participation.

To generate participation assume that a leverage factor of 4 applies, i.e. the investor does not participates with CHF 25 (the investment account) but in fact with CHF 100 at day 1. The difference of CHF 75 is money borrowed from the issuer at a borrowing rate, see Figure 3.16. Suppose that at day 2 the underlying value increases by 3 percent. The investment account has then the value CHF 28, debt financing is increased to CHF 84 to maintain the leverage factor constant. The invested amount raised to CHF 112, i.e. CHF 28 plus 84. This shows that an increase of the investment account without leverage, i.e. 3 percent of CHF 25 which is CHF 0.75 is leveraged with the factor 4 to an increase of CHF 3 which means a 12 percent increase. The figures at day 3 follow with the same logic if a constant leverage factor is restored.

We write I_t for the value of the investment account at time t. The dynamics of this account is given by

$$\begin{aligned} I_t &= I_{t-1} \\ &+ (I_{t-1}+F_t)\left(R_t - g\frac{n}{360}\right) \\ &- N \times \text{Fee} \times \frac{n}{360} \\ &- F_t(r_t + \text{DebtFee}_t) \times \frac{n}{360} \end{aligned} \tag{3.9}$$

where

- F_t the debt financing level.
- R_t the return of the underlying value.

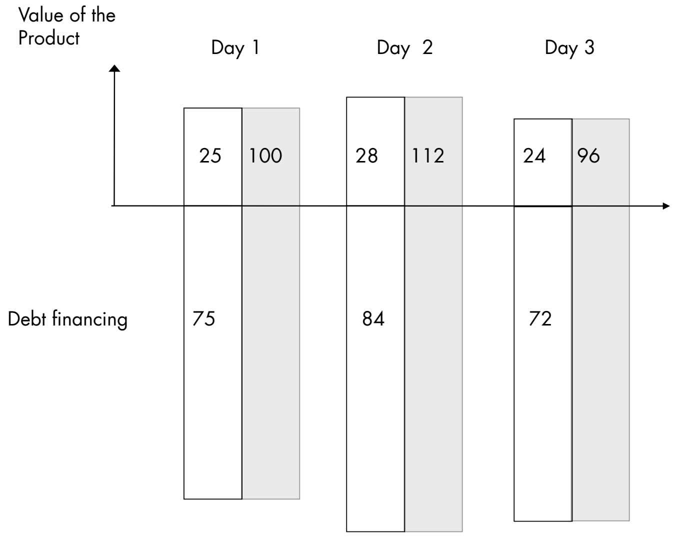


Figure 3.16: Leverage effect mechanics.

- g the gap risk fee.
- n the number of days between $t-1$ and t.
- r_t the interest rate SARON - Swiss Average Rate Over Night.
- DebtFee_t the add on fee for debt financing.

The intuition of the dynamics (3.9) is as follows. The new value of the investment account is equal to the past value plus or minus some returns or costs. The first return line $(I_{t-1}+F_t)\left(R_t - g\frac{n}{360}\right)$ expresses the growth approximatively of the old account value time the leverage factor h, i.e. $I_{t-1}+F_t \sim h \times I_{t-1}$, with the growth factor 'return minus gap risk fee costs' $(R_t - g\frac{n}{360})$. The expression $N \times \text{Fee} \times \frac{n}{360}$ is simply the management fee and the last line are the debt capital cost for the client.

The dynamics of the investment account is of first order and it can become negative. That is, if say the return becomes negative I_t can become very small such that a jump event can lead to $I_t < 0$. If this happens the value of the structured product is lower than the capital protection part. To avoid this, the issuer carries this so-called gap risk and the investor pays for this. If I_t falls below a certain level, I_t is set equal to zero and the investor then has a buy-and-hold product which consists of the capital protection part only.

So far the leverage factor h was kept constant. We can make this factor state dependent relying on the negative leverage effect: If volatility increases, the value of the corresponding risky asset decreases. Figure 3.17 shows this effect for the Swiss index SMI and the corresponding volatility index. The mirror behavior is almost perfect i.e. showing negative correlation.

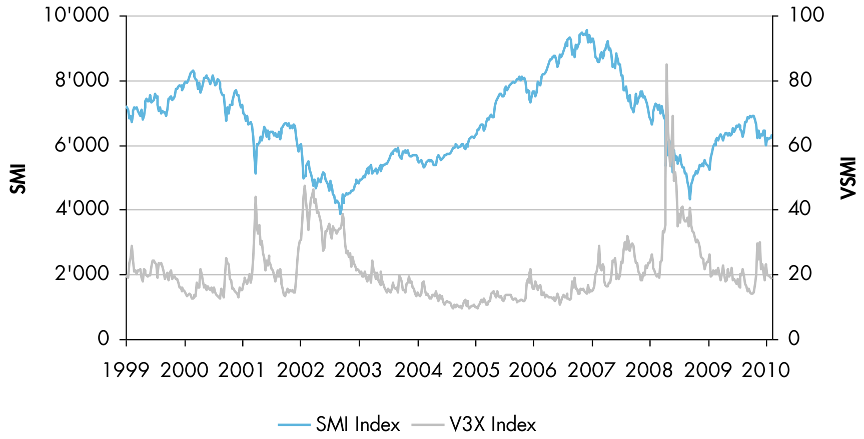


Figure 3.17: Negative leverage effect between SMI and VSMI.

We can use this in the CP as follows. Instead of working with a fixed leverage factor we make this factor volatility dependent. If volatility raises we expect risky assets to fall and therefore we decrease the investment in the risky investment account, i.e. the leverage factor decreases and contrary, if volatility increases. There are many possible ways how the leverage factor can be made volatility state dependent. In general $h_t = f(\bar{h}, \bar{\sigma}, \sigma_t)$ with $\bar{h}$ and $\bar{\sigma}$ fixed reference values, i.e. one assumes that for a given underlying there is a kind of equilibrium or normal model leverage factor and volatility. One could then model h_t as a mean reverting process or simpler setup a rule of thumb, such as

$$h_t = 2\frac{\bar{h}}{\sigma_t/\bar{\sigma} + 1} \ .$$

In any case, one has to specify the equilibrium values and time dependent volatility. The equilibrium volatility can be taken as the historical average volatility. The equilibrium leverage factor should be chosen with care since gap risk and level of leverage are intimately related to each other as we discuss below. Since the markets for these products do not possess options the time varying volatility σ_t can not be chosen to be an ATM implied volatility. One has to choose a time-weighted average of historical volatilities. If one chooses the weight to be present-oriented, i.e. for example the typical exponential time decay into the past, one can capture recent moves with a satisfactory precision.

Example

Figure 3.18 illustrates the discussion for a German stock exchange listed real estate fund over the last ten years. The real estate fund performed well until the start of the financial crisis (right panel). The CP where the investment account consists of leveraged investment in the fund does not suffers from the drop. The only effects are that the investment account becomes smaller (left panel) and that the investment leverage in this risky asset was reduced (right panel).

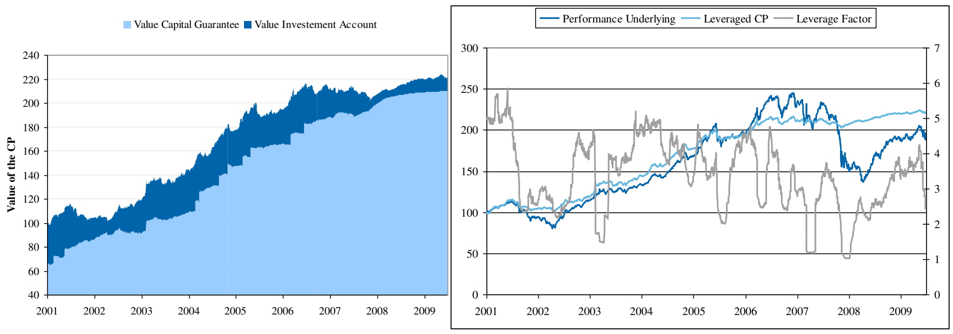


Figure 3.18: Performance of a listed German real estate fund (right panel), performance of the CP with investment into the real estate fund (right panel) and variable leverage factor over time (right panel). The left panel shows the CP value evolution where the two components capital protection and investment account are separately shown. *Source: Zurich Cantonal Bank.*

On the left panel the capital protection amount, which was set equal to 100 percent at the beginning is increasing over time in a jump style. This reflects the lock-in feature: If the real estate fund performs well and reaches a defined value than a fraction of this risky investment is transferred into the capital protection.

We considered so far the risky investment part I_t of the CP and neglected the capital protection part. Since we want to relate the discussed structure to the Constant Proportion Portfolio Insurance (CPPI) strategy we consider both components of the structured product in the sequel. We first recall the CPPI strategy. A CPPI strategy stats with an initial investment above the present value of a guarantee and stops immediately when the strategy value hits the present value of the guarantee. That is, a guarantee level can be reached with probability one as long as it is possible to stop exactly at the moment when the portfolio value is equal to the present value of the guarantee and this amount grows at the constant, risk free interest rate up to maturity. CPPI is a portfolio insurance strategy designed to achieve a minimum level of wealth while at the same time participating in upward moving markets.

We assume that there are two assets, a risky asset S_t which satisfies a geometric Brownian motion law with constant mean μ and constant volatility σ and risk free asset B_t with constant return r. At each time t a fraction a_t of the portfolio is invested in the risky asset and $1 - a_t$ in the risk free one. The portfolio value dynamics V_t is therefore given by

$$dV_t/V_t = a_t dS_t/S_t + (1 - a_t)dB_t/B_t \ , \ V_0 = x \ .$$

This strategy is model independent i.e. it suffices to observe the asset prices but it is not necessary to know the distribution of the prices. For a given time horizon T, the capital guarantee is G at time T and its present value is denoted $F_t = e^{-r(T-t)}G$. At each time t the difference $V_t - F_t = c_t$ is called the cushion. Leverage is introduced by a positive factor h, i.e. the investment in a_t becomes

$$a_t := \frac{hc_t}{V_t} = h(1 - \frac{F_t}{V_t}).$$

The following proposition summarizes the CPPI facts.

Proposition 3.4.1. *Assume that S satisfies a geometric Brownian motion dynamics. The cushion process c_t is lognormal, i.e.*

$$dc_t/c_t = ((r + h(\mu - r))dt + \sigma h dW_t) \ .$$

The value of V_t is given by

$$V_t = \underbrace{Ge^{-r\tau}}_{PV\ Guarantee} + \underbrace{\frac{V_0 - Ge^{-rT}}{S_0^h} e^{\left(r - h(r - \frac{1}{2}\sigma^2) - h^2\frac{\sigma^2}{2}\right)t} S_t^h}_{Power\ Option} \ . \tag{3.10}$$

The proof is given in Appendix 3.4.1. Thus, the value process of a simple CPPI strategy is path independent. For $h > 1$ the payoff is convex, i.e. a power option. Setting $t = T$ is follows that with certainty the terminal value of the strategy is higher than the guarantee. Although the CPPI strategy is model independent, the evaluation of the strategy at future date, the calculation of the final performance or risk figures require the specification of a model.

Given the distribution of the value process, the expected value and the variance follows at once:

$$E[V_t] = F_t + (V_0 - F_0)e^{(r+h(\mu-r))t}$$

and

$$\mathrm{var}[V_t] = (V_0 - F_0)^2 e^{2(r+h(\mu-r))t}\left(e^{h^2\sigma^2 t} - 1\right) .$$

How is CPPI related to the former CP structured note discussion? Consider the dynamics of the investment account I_t in (3.9). If we neglect the fee structure, assume a constant leverage factor, set gap risk equal to zero, using $I_{t-1} + F_t \sim h \times I_{t-1}$ and assume constant interest rates, then the dynamics is given by

$$\frac{I_t - I_{t-1}}{I_{t-1}} = h\Delta R_t - (h-1)r\Delta t .$$

Setting $\Delta R_t = \Delta S_t / S_t$ we get

$$\frac{\Delta I_t}{I_t} = h\Delta S_t/S_t - (h-1)r\Delta t .$$

Comparing this with the c-dynamics

$$dc_t/c_t = hdS_t/S_t - (h-1)rdt$$

it follows that the I_t-investment account dynamics is under the assumptions equal to the cushion dynamics of the simple CPPI.

In the above discussion we got the identification of the CPPI strategy with the investment account dynamics of the CP under several assumptions. Most of them are neither important in practice nor do they provide any interesting insight. The exception is the gap risk. First of all, this risk exists naturally and second, one should not issue or sell products with risks which one cannot accept or reject simply because one has no idea about the severity of the risk. We consider this risk in more detail.

Gap risk is related to jumps in the risky asset value given a low cushion or investment account. Instead of introducing a difficult to calibrate jump process the risk is introduced by restricting trading to discrete dates $t_0 < t_1 < \ldots < t_n = T$, which we assume to be equidistant. But the underlying process trades at much higher frequency, i.e. we assume continuous trading of the underlying. This approach makes perfect sense from a practitioners view where rebalancing between the leveraged risky asset investment and the capital guarantee part takes place only weekly or monthly depending on the volatility and liquidity of the underlying instrument. The restriction that trading is only possible immediately after any discrete date implies that the number of shares held in the risky asset is constant on the intervals between two trading dates. But the fraction of wealth invested in the risky asset changes within a given period due to the risky asset price fluctuations. We let

$$\psi_t = \frac{a_t V_t}{S_t} = \frac{hc_t}{S_t} , \ \phi_t = \frac{(1-a_t)V_t}{B_t}$$

denote the respective investments into the risky asset and the risk free one and $\xi_t = (\phi_t, \psi_t)$ is the CPPI portfolio or investment strategy in continuous time. We discretize this strategy. Suppose that the discretization ξ_t^n is defined by setting ξ_t^n equal to the left-end value ξ_{t_k} in a trading interval $(t_k, t_{k+1}]$ and keeping the strategy fixed for the whole interval time, then the discretized strategy will not be self-financing: The strategy is kept constant whereas asset values can change. One obtains a self-financing strategy if the risk free investment is set equal to

$$\phi_t^n = \frac{1}{B_{t_k}^n}\left(V_{t_k}^n - \psi_t^n S_{t_k}\right) \ , \ t \in t_k, t_{k+1}]$$

and the discrete risky investment - avoiding short positions in the risky asset - is

$$\psi_t^n = \max\left(0, \frac{hc_{t_k}^n}{S_{t_k}}\right) \ .$$

However, similar as for the simple continuous time CPPI, the discrete time CPPI does not include short sale restrictions on the risk less asset. Recall that constant proportion portfolio insurance means that the fraction of wealth á which is invested in the risky asset is given proportionally to the difference of the portfolio value and the floor, i.e. the cushion. To estimate the probability that the cushion becomes negative, i.e. gap risk is realized, we consider the cushion process, i.e. the difference

$$c_{t_{k+1}}^n = V_{t_{k+1}}^n - F_{t_{k+1}}.$$

This difference can be explicitly calculated, see Appendix 7.36. The result consists of two cases. If the first time where cushion is negative after maturity date, the value of the strategy is the discrete time analogue of (3.10). If cushion is negative before maturity, the value at this date is pushed forward with the risk free rate to maturity date.

Using the results stated in the Appendix of [?], we define the following risk measures for gap risk and prove the stated results.

Proposition 3.4.2. *Assume the above discrete/continuous time model. Local shortfall probability P_i^{loc} is given by*

$$P_i^{loc} := P\left(V_{t_{i+1}}^n \leq F_{t_{i+1}} | \ V_{t_i}^n > F_{t_i}\right) = P^{loc} = \Phi(-d_2) \tag{3.11}$$

with

$$d_2 = \frac{\ln(h/(h-1)) + (\mu - r)T/n - \frac{1}{2}\sigma^2 T/n}{\sigma\sqrt{T/n}}$$

d_2 is independent of the date t_i. The shortfall probability P^{short} that the cushion is not positive at maturity T is given by

$$P^{short} = 1 - (1 - P^{loc})^n \ . \tag{3.12}$$

Expected shortfall $ES := E[G - V_T^n | V_T^n \leq G]$ *is given by*

$$ES = -\frac{(V_0 - F_0)k_2 \frac{e^{rT} - k_1^n}{e^{rT/n} - k_1}}{P^{short}} \tag{3.13}$$

where

$$k_1 = he^{\mu T/n}\Phi(d_1) - e^{rT/n}(h-1)\Phi(d_2) \ , \ k_2 = he^{\mu T/n}\Phi(-d_1) - e^{rT/n}(h-1)\Phi(-d_2)$$

and $d_1 = d_2 + \sigma\sqrt{T/n}$.

It follows that if trading becomes continuous ($n \to \infty$) the shortfall probability P^{short} converges to zero. One might guess that the shortfall probability is monotonically decreasing in the hedging frequency n. In general, this is only true after a sufficiently high n is reached. The effect that the shortfall probability is increasing for small n is more pronounced for high volatilities and high leverage factors. In contrast to a discrete time option based strategy with a synthetic put, the calculation of the shortfall probability implied by a CPPI strategy is simple. This follows if one observes that the shortfall event is equivalent to the event that the stopping time is prior to maturity.

Figure 3.19 shows the behavior of the different risk measures when strategy parameters such as the leverage factor or the guarantee level are changed or when model parameter such as the volatility, the drift or the interest rate vary.

Sensitivity calculation and the figure in 3.19 show that:

- Shortfall probability increases if leverage increases, drift decreases or volatility increases. The shortfall level is independent of the guarantee level.
- Expected shortfall is increasing in an increasing drift, volatility or leverage factor. An increasing guarantee level reduces expected shortfall.
- An increasing trading frequency lower all risk figures.
- Increasing maturity and increasing the trading frequency such that the ratio remains constant and adjusting the initial guarantee value due to the increased trading frequency leads to an increase of shortfall probability and expected shortfall.

The CPPI strategy considered implies that if once the value of the strategy hits the present value of the floor the investment in the risky asset is zero and remains zero with probability one until maturity. One might think to alter the CPPI strategy such that this unpleasant investment effect does not follow. Basically the change is to save the value of the risky investment, i.e. to lock-in this value into cash, if the risky investment approaches the value zero before the event of zero investment in the risky asset is met. One defines such a risk measure, the cash-lock probability, by the conditional probability that given at a date the investment in the risky asset is yet low as the probability that in a future date the investment amount in the risky asset will even be lower than a defined threshold. Given such a risk measure and a continuous time CPPI cash-lock probability

Leverage Factor	12	12	12	12	12	12	**5**	**5**	**5**
Volatility	**0.1**	**0.2**	**0.3**	**0.1**	**0.2**	**0.3**	**0.1**	**0.2**	**0.3**
Risk Free Rate	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Drift Underlying	0.085	0.085	0.085	0.085	0.085	0.085	0.085	0.085	0.085
Maturity	1	1	1	1	1	1	1	1	1
Trading Frequency	12	12	12	**52**	**52**	**52**	12	12	12
Initial Guarantee Value	95	95	95	95	95	95	95	95	95
Initial Product Value	100	100	100	100	100	100	100	100	100
Factor for Expected Shortfall	-0.001	-0.289	-1.864	-0.000	-0.005	-0.222	0.000	-0.000	-0.008
Factor for Expected Shortfall k_1	1.039	1.056	1.117	1.009	1.009	1.012	1.019	1.019	1.019
Factor for Expected Shortfall k_2	-0.000	-0.017	-0.077	-0.000	-0.000	-0.003	0.000	-0.000	-0.001
d_2	3.101	1.529	0.995	6.316	3.148	2.087	7.817	3.887	2.567
Local Shortfall Probability	0.1%	6.3%	16.0%	0.0%	0.1%	1.8%	0.00%	0.01%	0.51%
Shortfall Probability	1.2%	54.3%	87.6%	0.0%	4.2%	62.0%	0.00%	0.06%	5.98%
Expected Shortfall	0.582	2.809	11.879	0.148	0.555	1.814	-0.068	0.305	0.642

Figure 3.19: Risk measures under simple CPPI.

is one if conditioning at a given date is one zero-investment and it is given by a normal distribution in all other cases. It follows that the cash-lock probability is not monotone in the time to maturity level and the multiplier level.

So far gap risk was considered without jumps. We consider the case where we believe that gap risk is due to jumps. That is we consider gap risk and its fair pricing and apply it to leveraged Delta one bond investments. This investment is issued as a Retail Structured Product (RSP). In this case Figure 3.20 illustrates this risk source.

It follows that the maximum loss possible for the issuer is the total lend capital for leveraging. The investor pays an annual fee to the issuer to compensate him for the gap risk.

We assume that jumps in the price process of the bond portfolio which cause a 'gap event' are exclusively generated by a credit event. This is plausible since jumps associated with market risk, i.e. moves in the interest rate, will in general be small and hence their contribution for the gap risk will be small as well. Ie essentially 'tail risk' from credit will be taken into account. Liquidity risk which might also cause a gap event can be assumed to be absorbed into parameter α introduced below. To estimate the gap risk generated by credit risk / events we use a hazard rate model:

- The probability that a gap event with a Loss L occurs in $[t, t+\delta t]$ is $\lambda_t^L \delta t$ conditional

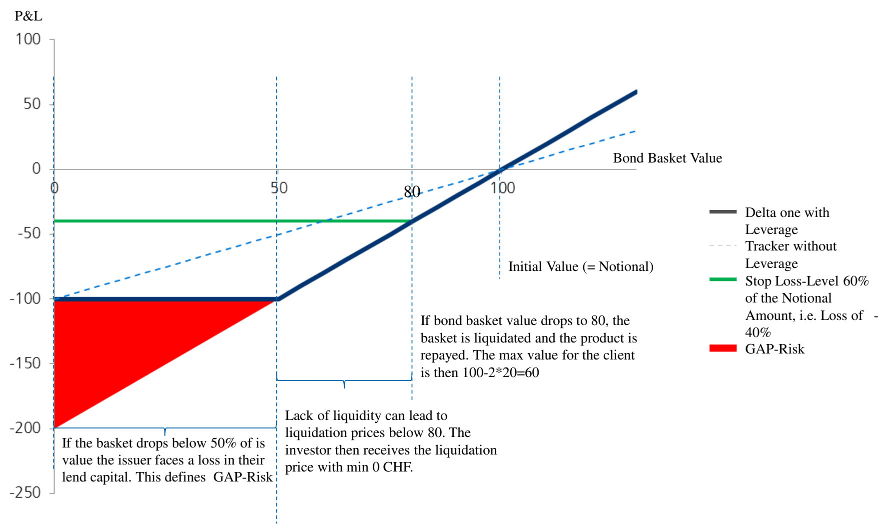


Figure 3.20: Gap risk for a leveraged Delta 1 bond investment with leverage factor 2.

that none has occurred yet.

- The hazard rate λ_t^L itself is assumed to have an exponential distribution (Pareto distribution for peak over threshold):

$$\lambda_t^L = \exp(-\alpha L)$$

Using these assumption we calculate the gap fee g in an interval $[0, T]$:

$$\begin{aligned} g &= \frac{1}{T}\mathbb{E}[D_\tau L_\tau \theta(T-\tau)|\mathcal{F}_0] + \mathcal{O}^2(r,\lambda) \\ &= \frac{1}{T}\int_0^T \int_0^\infty \exp\big(-(r_s+\lambda_s^L)\big)\lambda_s^L L dL ds + \mathcal{O}^2(r,\lambda) \\ &= \frac{1}{T}\int_0^T \int_0^\infty \exp(-\alpha L) L dL ds + \mathcal{O}^2(r,\lambda) \\ &= -\frac{1}{\alpha}\exp(-\alpha L)L\Big|_0^\infty + \int_0^\infty \exp(-\alpha L) dL + \mathcal{O}^2(r,\lambda) \\ &= \frac{1}{\alpha} + \mathcal{O}^2(r,\lambda)\ . \end{aligned}$$

We need to estimate the analogue of α for a single bond which we call β. To do this we assume that jumps are again only generated by credit events or due to sudden widening of the bid / ask spread (liquidity risk). Both are assumed to be priced into the credit spread of the bond. The hazard Rate of the bond and distribution of jumps / losses are modelled in the same way as above which implies for $\lambda_t^L = \exp(-\beta L)$:

$$c_T = \frac{1}{T}\ln\big(\frac{\int_0^T\int_0^\infty \exp\big(-(r_s+\exp(-\beta L))\big)dLds}{\int_0^T \exp(-r_s)ds}\big) \tag{3.14}$$

$$= \frac{1}{\beta} \tag{3.15}$$

We next derive the relation between α and β. The probability of jumps crossing the threshold S reads $p_{Hit} = \frac{L-1+S}{L}$ with L the leverage factor and S the threshold relative to notional amount of the RSP. Hence,

$$\Rightarrow \alpha \sim \frac{L}{L-1+S}\beta\ .$$

Since $\frac{1}{\beta}$ is a measure for the magnitude of the jump for a single bond, $\frac{1}{\alpha}$ is a measure for the jump below a threshold S:

$$\frac{1}{\beta} \sim \quad V_{t-} - V_{t+}\ ,\ \frac{1}{\alpha} \sim \quad S - V_{t+}\ . \tag{3.16}$$

Assuming

$$\mathbb{E}[S] \simeq \frac{V_{t-}+V_{t+}}{2}\ ,$$

we get:

$$\Rightarrow \alpha \sim 2\beta.$$

Finally, for the intensity $\lambda \sim \frac{1}{\alpha}$ of the price process of the bond basket. If we assume N bonds, $\overline{\beta}$ the average β of a bond and ρ the collective correlation parameter of intensity processes we obtain:

$$\Rightarrow \alpha \sim \frac{\overline{\beta}}{\rho + \frac{1-\rho}{N}} = \frac{1}{\overline{c}(\rho+\frac{1-\rho}{N})}$$

Taking all multiplication factors into account[9] we finally obtain:

$$g \simeq \frac{1}{\alpha} \simeq \frac{\overline{c}(\rho+\frac{1-\rho}{N})(L-1+S)}{2L}$$

As an example, we assume that the bond basket consist of 5 AA or A bonds. The average credit spread is $\overline{c} \in [6, 67]$ bps for AA bonds and $\overline{c} \in [13, 128]$ bps for A bonds. Leverage is $L = 2, \rho = 0.5$ and the threshold $S = 0.6$. This implies the following prices for gap risk:

[9] $\overline{c}$ is the average credit spread of bond basket.

$$
\begin{aligned}
g_{AA} &\in [1, 16] \text{ bps} \\
g_A &\in [3, 32] \text{ bps}
\end{aligned}
$$

With the conservative estimate $\rho = 1$ we get:

$$
\begin{aligned}
g_{AA}^{\rho=1} &\in [2, 30] \text{ bps} \\
g_A^{\rho=1} &\in [5, 58] \text{ bps}
\end{aligned}
$$

3.5 View and Trade

View and trade is another method of investment which in some sense is the opposite to passive investment. The investor has a clear view and wants to implement the view efficient, effective and transparent in a trade. The investor is not driven by diversification motivation. Contrary, often singular events create opportunities which the investor wants to realize.

3.5.1 Butterfly

A **butterfly** trade, is based on an investor which believes that the underlying value will be close to its today value and which wants to make a profit out of this believe. If the belief turns out to be wrong, only small losses follows. A butterfly is long a **in-the-money (ITM)** call , long a **out-of-the-money (OTM)** call and short two **at-the-money (ATM)** calls, see Figure 3.21. The different strikes are ordered as $K_L < K_M < K_H$.

An investor observes three call options on the same underlying value with the following strikes and prices:

- Call with strike 45, price 8.
- Call with strike 50, price 4.
- Call with strike 55, price 1.

The investor beliefs that the underlying will not move. A butterfly strategy allows him to implement his view. He buys the 45 and 55 call and sells the 50-call.

3.5.2 Leveraged Negative Basis

The butterfly trade was a simple example. The trade under consideration is more complex. The **basis** in credit risk is the difference in the valuation of credit risk of a counter party once in the **bond market** and once in the **CDS market**. More precisely the Asset Swap Spread (ASW) is defined for a bond as the solution of

$$\text{Bond price} = \sum_{j=1}^{N} \frac{1}{(1 + r_f + \text{ASW})^j} c_j$$

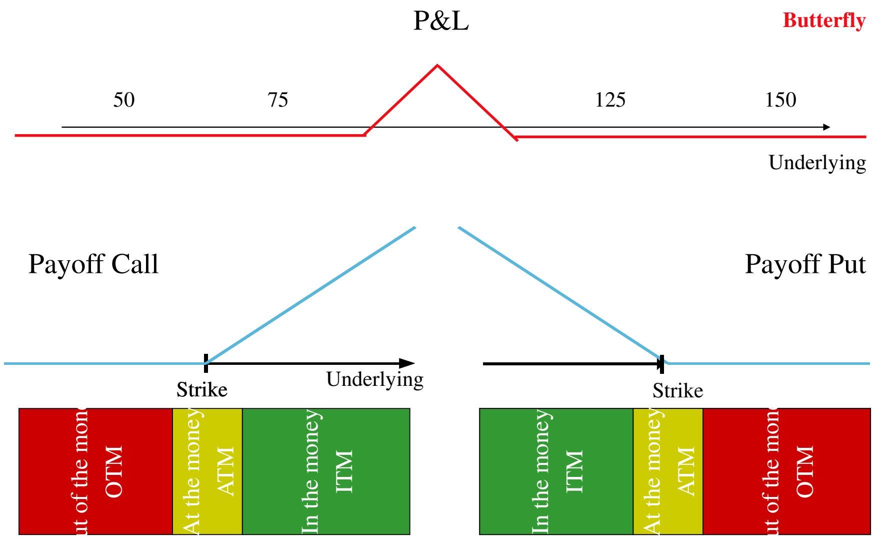


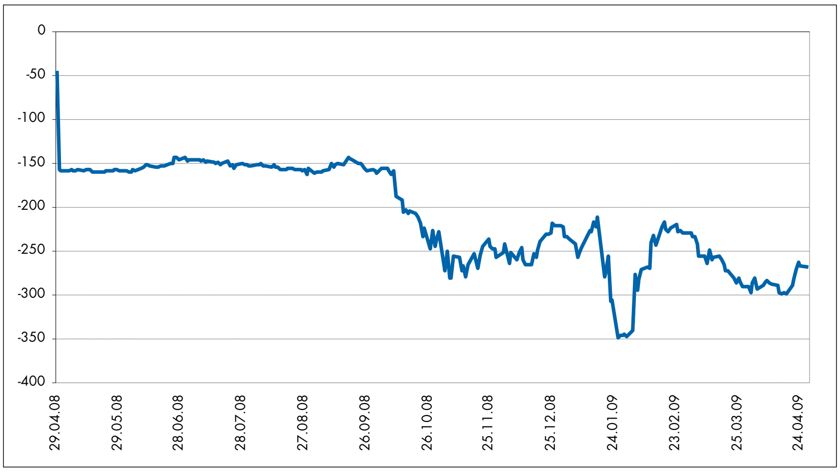
Figure 3.21: Butterfly.

with r_f the risk free rate. If we denote by CDS the CDS premium the difference

$$B = \text{CDS} - \text{ASW}$$

is called the *basis*. This difference is usually close to zero. This reflects that the price of credit risk is valued the same in both markets. But in the financial crisis the difference became strongly negative for a large number of corporates, see Figure 3.22.

The figure shows that the difference became substantially negative and that this was a persistent fact. For this securities the cost to buy protection via the CDS was much cheaper than the drop in bond prices due to the creditworthiness of the bond issuers under stress. There are many possible explanations for this observation. A major role is liquidity which became the key risk factor in the financial crisis. Since one has to pay the notional amount in the bond market but only the risk premium CDS in the credit derivative market the rise of the ASW compared to the CDS premium can be interpreted as a liquidity trigger. The price shock is mostly due by large international banks which had to shorten their balance sheet considerably in the financial crisis. That for assets from the debt sector had to be massively sold. At the end of 2009 the value of all assets in the Swiss banking sector was seven time the GDP. Before the crisis the difference was given by a factor of nine. Given the importance and activity of the two big banks UBS and CSG this reduction in the factor gives an estimate of the reduction of assets. A different structural reason for a negative basis are funding costs above LIBOR for traders or shorting corporate bonds means that one has to borrow the bond at a repo rate which



Source: Bloomberg

Figure 3.22: Negative Basis for six bonds during the financial crisis 2008-2009. *Source: Bloomberg*

is often illiquid and therefore higher than the CDS premium. In the years before the financial crisis the basis was slightly positive (10-20bps) on average. Such a situation occurs if (i) bonds are trading below par such that the loss of a given bond default is smaller than the CDS payoff, (ii) LIBOR is not risk free and therefore the spreads for very good rated entities are low or even negative and (iii) corporate bonds may have important change clauses.

This wide negative basis can be used to structure an investment product as follows: Basically if we can lock in the negative basis for a given period each day where the negative basis exists we will obtain a cash flow: That is we buy a portfolio of bonds and the corresponding CDS to hedge credit risk where a the two portfolio generate a significant negative basis. We don't know how long the basis remains negative, i.e. when will the significant basis vanish and the product will be terminated by unwinding the bonds and CDS. One could also think to structure a similar trade for a positive basis. But in this case one has to go short the instruments which is much less efficient than going them long.

Suppose that we setup a portfolio of bonds and CDS such that to each bond corresponds a CDS where the reference entity is identical to the bond issuer. Assume that the basis is significantly negative for each bond-CDS pair, say for 200 bps. Suppose that the issuer of a structured note is either cash rich or can borrow at a low rate. Then the 200 bps are leveraged, by a factor of 5 for example. Then, we have a gross coupon of 10 percent which can be paid annually to the investor. From these 10 percent one deduces the cost of financing the leverage and the management fee. The risk for the investor is the joint cross-default of a bond and the corresponding CDS - an event which has a very low probability to occur. But if this happens, most of the capital is at risk. Suppose that the leverage is 10 and the recovery rate is 10 percent for a bond. Then the lost capital will be by a single joint cross-default in the whole portfolio

$$9 \times 10\% = 90\%$$

of the invested capital.

The structuring is illustrated in Figure 3.23. The issuer invests 100 in a currency (here Euro). This investment is leveraged with a factor $h = 10$, i.e. 900 are borrowed from the treasury. With the total amount of 1'000 the bond portfolio is bought (long) and the CDS portfolio with the corresponding notional amounts and reference entities is setup. The return from the portfolios are interest rate earnings, here 12m EURIBOR, and the negative basis. 90 percent of interest rate earnings are used to compensate the treasury plus a fee of 45 bps. The remaining 10 percent minus the 45 bps go to the investor. The earnings from the negative basis are also split between the investor and the issuer. Typically, a high percentage say 75 percent goes to the investor and 25 percent to the issuer.

This rough structure is presented in more details next. A portfolio of bonds denominated in a currency and a portfolio of CDS in the same currency are setup. There is a cash portfolio which consists of the financing amount in a currency and the proceeds of the bond and CDS portfolio. The Underlying consists of a bond, CDS and cash portfolio at all times. The leveraged bond and CDS portfolios are marked-to-market. A leverage unit equals a fixed amount liquidity, which can be theoretically borrowed by the certificate holder to finance the additional units of the Bond portfolio in order to generate the leverage. At initial fixing, the cash portfolio only consists of the loan, i.e. (h - 1)$\times$ the Leverage Unit, in which the Leverage Unit figure is a negative number. Through the lifetime of the product, four additional parts accrue to the cash portfolio:

- Cash Part incoming from the Bond portfolio, which are aggregated Coupon payments and Capital redemptions of the Bonds, which have not been subject to a default or Adjustment Event.
- Cash Part relating to the CDS portfolio, which are premium payments on the CDS contracts and incoming payments from defaulted bonds.
- Cash Trigger Part, which are payments arising from Adjustment Events. Such an Event occurs if the Spread + negative Basis for one of the bonds is > 0 (or if the Calculation Agent acting in good faith and commercial reasonable matter determines at any time between the Issue Date and Redemption Date that an Adjustment Event has occurred), in which case a profit taking is initiated.

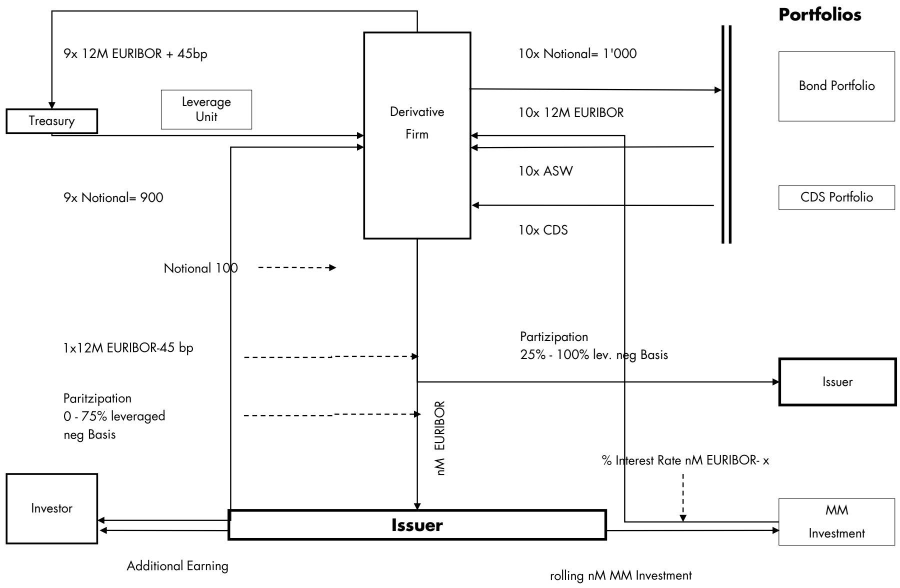


Figure 3.23: Structuring of a leveraged negative basis trade.

- Cash Certificate Part. The Certificate cannot pay out negative Coupons. Those are instead booked into the cash portfolio as expense surplus and charged on the following Coupon Payment. The Coupon Payments for every period are equal to the difference between the Return and the Financing Amount. The Return consists of the Interest Rate Reference and the negative Basis, whereas the Financing Amount for the period in question is enlarged with the Management Fee and a possible expense surplus from the preceding time period.

We consider an example, see Figure 3.24, for a leveraged negative basis transaction with capital protection. We start with a assumed negative basis of 120 basis points (280-400). The certificate has a 5 year maturity and we consider three scenarios: S1, the negative basis remains at 120 bps for all dates, S2 the negative basis decreases over time which is the scenario where the financial markets normalize and S3 the negative basis widens, i.e. the financial crisis starting in 2008 will be followed by say a double dip. If the basis remains constant, there is no impact on the certificate value, i.e. the product trades always at 100 percent. In all scenarios the final value of the product is 100, i.e. the protected capital. If the negative basis decreases over time the value of the certificate will increase over 100. The value of 114 in S2 is calculated as follows:

$$114 = (\text{NegBas at 0 -NegBas at } t) \times \text{BPV(t)} \times \text{LevFac} + \text{NegBas at 0}\ .$$

The earnings from the negative basis in almost all scenarios equal to 900 bp per annum. This value is calculated as follow. If the trigger value is not reached, i.e. if the value of the certificate at a given date divided by the initial certificate value is not below 0.3 then a positive return follows. In the other case, the coupon is zero which is the case in scenario S3 where the basis raises very strong in year 2. The positive return of 900 bps p.a. is equal to the negative basis at time 0 times the participation rate - here 75 percent - times the leverage factor, here $h = 10$.

	0y	1y	2y	3y	4y	5y
Spread Bonds [bps]	400					
Spread CDS [bps]	280					
Negative Basis Scenarios						
Neg. Basis S1 [bps]	120	120	120	120	120	120
Neg. Basis S2 [bps]	120	96	72	48	24	0
Neg. Basis S3 [bps]	120	350	240	350	800	120
Market Value of Negative Basis Contribution to the Certificate						
MtM S1 [%]	100	100	100	100	100	100
MtM S2 [%]	100	109	114	114	109	100
MtM S3 [%]	100	13	65	55	33	100
Negative Basis Return						
Negative Basis Return	900	900	900	900	900	900
Negative Basis Return	900	900	900	900	900	900
Negative Basis Return	900	0	900	900	900	900

Figure 3.24: Illustration of some negatives basis trade aspect.

We refer to the Appendix 7.8.2 for a quantification of the different return and cost components. We only consider on component in the main text - the coupon part. First, coupons can never be negative. At all times, coupons consist of four parts: The return for the period from the certificate, the interest on the financing amount for the period, the fixed management fee and possible expense surplus from the preceding time period: If during an interest rate period expenses (Financing and Fee) surmount the returns, the result is a theoretical negative coupon, which will not be paid out. This expense surplus will instead be charged on the following interest rate period. At a default event, the CDS portfolio contribution is at market value. The cash portfolio consists of 3 further components: The loan part, the coupon part of the certificate and the trigger part in case of a trigger event. We consider the coupon part. Coupons can never be negative. At all times, coupons consist of four parts: The return for the period from the certificate, the interest on the financing amount for the period, the fixed management fee and possible expense surplus from the preceding time period: If during an interest rate period expenses (Financing and Fee) surmount the returns, the result is a theoretical negative coupon, which will not be paid out. This expense surplus will instead be charged on the following interest rate period. If the coupon of the last interest rate period is negative, the resulting expense surplus will be subtracted from the redemption amount of the certificate. The coupon part of the certificate is defined as follows:

$$\begin{aligned} V^{\text{Coupon Cert}}(t) &= \sum_p \chi_{\{\tilde{t}_p<t\}} N \\ &\times \max\left(\mathcal{E}(p) + h \times \mathbf{p} \times R \times \bar{\mathbf{n}} - (h-1)F(p) - G + s(p-1), 0\right) \end{aligned}$$

where:

$$\begin{aligned} \mathcal{E}(p) + h \times \mathbf{p} \times R \times \bar{\mathbf{n}} &= \text{Earnings from the Negative Basis} \\ -(h-1)F(p) - G + s(p-1) &= \text{Total Costs} \end{aligned}$$

where $\mathcal{E}(p)$ are earnings from the interest rate reference (12m EURIBOR for example) and $h \times \mathbf{p} \times R \times \bar{\mathbf{n}}$ is the return from the negative basis. The negative basis is fixed at initial fixing as follows: For each Bond issuer, the basis is calculated, the weight for each Bond relative to the entire bond portfolio is calculated an the sum of the weighted negative basis for all bonds constitutes the negative basis at initial fixing. The negative basis at initial fixing is locked in for the return calculations, except the weight of a bond is set to 0 because of maturity, sale or default. The negative basis contributes every day a positive amount to the return, if the trigger event has not been activated, or zero, if the trigger event has been activated. The daily trigger event is activated if the calculation agent observes the value of the certificate below a fixed barrier level - the trigger ratio R - of the certificate denomination. If the trigger event is not activated, the investor participates to the degree of $\mathbf{p}$ on the negative basis or any changes caused by reaching the maturity, sale or default of the bonds.

If expenses exceed returns in one period, the following mechanism applies where a balance account s is introduced. The balance $s(p)$ in period p satisfies the recursion relation:

$$\begin{aligned} s(1) &= 0 \\ s(p) &= \min\left(E(p) - F(p) - G + s(p-1), 0\right) \ , \ p = 2, 3, 4, 5 \ . \end{aligned}$$

The participation part from the coupon part of the certificate belonging to the issuer - i.e. the $(1-\mathbf{p})/\mathbf{p}$-fraction of the earnings from the negative basis - is denoted

$$V^{\text{Coupon Cert Issue}}(t) \ .$$

The contribution from the trigger event reads:

$$V^{\text{Trigger}}(t) = h \sum_{j=1}^{n} \chi_{\{t_T^j < \min(t_D^j, t)\}} w^j \left(V^{j,\text{Bond}}(t_T^j) - V^{j,\text{CDS}}(t_T^j)\right) \overline{D}(t_T^j, t),$$

i.e. only counter parties contribute to this part of the portfolio if trigger date is before default and maturity date of the counter party. The contribution size is then the difference between the bond and value of the corresponding counterparty at trigger date. The portfolio contribution of the trigger part is always positive, i.e. profit taking follows in such an event.

The loan for one leverage unit is:

$$V^{\text{Loan}}(t) = \text{LE}(h-1)\overline{D}(0,t)^{-1} \ , \ LE \ < \ 0 \ .$$

In summary, the cash portfolio reads:

$$\begin{aligned} V^{\text{Cash}}(t) &= V^{\text{Bond, Cash}}(t) + V^{\text{CDS, Cash}}(t) + V^{\text{Trigger}}(t) - V^{\text{Coupon Cert}}(t) \\ &- V^{\text{Coupon Cert Issue}}(t) + V^{\text{Loan}}(t) \end{aligned}$$

At $t = 0$, the Cash portfolio equals the loan value.

3.6 Fund Investments

3.6.1 Overview

Funds (mutual funds, investment funds) are **collective** investment schemes which pool money from investors. Funds can be classified from different views:

- **Open- or closed-end funds**. Open end funds are forced to buy back fund shares at the end of every business day to at the net asset value (NAV). NAV equals the current market value of a fund's holdings minus the fund's liabilities. Prices of traded shares during the day are also in term of the NAV. Total investment vary based on share purchases, share redemptions and fluctuation in market valuation. There is no limit on the number of shares that can be issued. **Closed-end funds** issue shares only once. Shares are listed on a stock exchange and trading occurs via the exchange: An investor cannot give back his shares to the fund but he has to sell them to another investor in the market. Prices of traded share can be different to the NAV, i.e. either they are higher (premium case) or lower (discount case). The vast majority of funds are of the open-end style. The worldwide investment fund assets in Q3 2011 were 18.6 Trillion Euros (Source: EFAMA).

Feature	**Open-end**	**Closed-end**
Number outstanding shares	flexible	fixed
Pricing	daily NAV	Continuous demand and supply
Redemption	at NAV	via exchange
Market Share	>95 percent	< 5 percent
US terminology	Mutual fund	Closed end fund
UK terminology	Unit trust	Investment trust
Cont. European (incl. CH) terminology	SICAV	SICAF

Table 3.4: Open-end and closed-end funds.

- **Exchange-traded fund (ETF)** are a mixture of open- and closed-end funds: They are traded at an exchange but the prices are close to the NAV by issuing.

- **Active vs. passive**. Costs for active management are higher than for passive one. Funds which replicate an index are called index funds. These funds are mostly considered to be passive since the replication of an index only requires a computer model with a modest logic and no human input in decision making. ETFs are often used to replicate indices. The replication of an index can be done in different ways. Consider a stock index. One can physically replicate the index by investing in the required number of stocks. The index can also be replicated using futures instead of the stocks. Wells Fargo first offered a not successful privately placed equally-weighted S&P 500 index fund in 1971 and a successful value-weighted one in 1973. It required Wells Fargo

  - to navigate regulatory and tax issues
  - surmount systems processing requirements
  - and educate potential investors.

  The introduction of index funds allows investors to complete the market at much lower transaction costs. The innovation was shaped by new technologies, was a response to tax and regulatory factors, and was driven by the presence of information asymmetries and transaction costs that made trading costly. The first fund was set up 1774 in the Netherlands. The fund invested in foreign government bonds, banks and West Indian plantations. The notion of diversification was spelled out explicitly in the prospectus of this first fund.

- **Investment classification.** Funds are classified according to their principal investments. The largest ones are money market funds, bond or fixed income funds, stock or equity funds and hybrid funds. Figure 3.25 shows the distribution the worldwide assets in the asset classes.

- **Fees and share classes.** Running a mutual fund involves costs, including shareholder transaction costs, investment advisory fees, and marketing and distribution expenses. All of the shares classes invest in the same portfolio of securities, but each has different expenses and, therefore, a different net asset value and different performance results. An important figure is the **Total Expense Ratio (TER)**. This is a percentage ratio defined as the ratio between total business expenses and the average net fund value. TER expresses the total of costs and fees which are continuously charged. Business expenses are fees for the fund's board of directors, the asset manager, the custodian bank, the administration, the distribution, marketing, calculation agent, audit, legal and tax authorities. Besides fees the calculation the performance is key. The following approach is widely used. Consider a period $[0, T]$.

$$\text{Performance }\% = \left(\frac{\text{NAV}_T \times f_1 \times f_2 \times \ldots \times f_n}{\text{NAV}_0} - 1 \right) \times 100 \,. \tag{3.17}$$

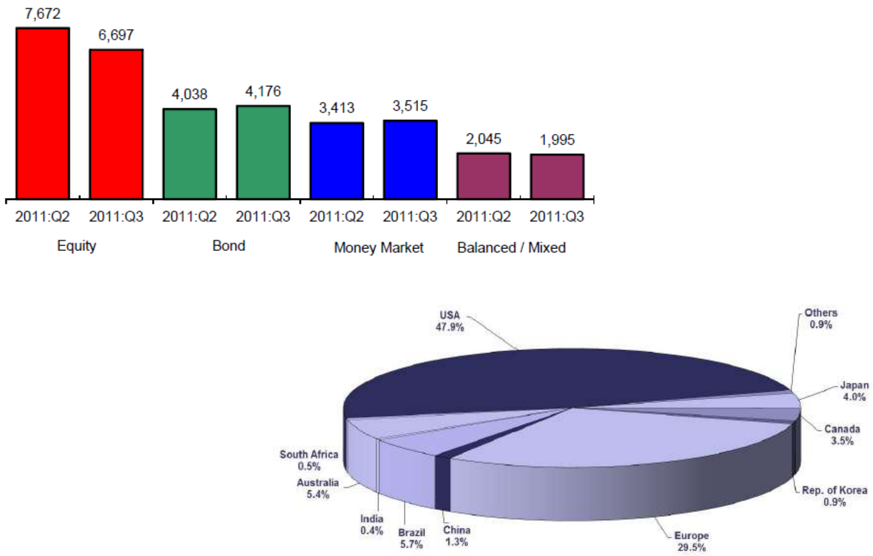


Figure 3.25: Left Panel: Worldwide Assets of Equity, Bond, Money Market, and Balanced / Mixed Funds. Right Panel: Distribution of investment fund assets.*Source: EFAMA International Statistical Release (2011:Q3), 2012*

f_j are the adjustment factors for the payout (dividends for example)

$$f_k = \frac{\text{NAV}_{ex} + BA}{\text{NAV}_{ex}}$$

with BA the gross payout, i.e. the gross amount of the earning- and capital gain payout per unit share to the investors and NAV_{ex} the NAV after the payout. Consider a NAV at year end 2005 of 500 Mio. CHF. 2006 earnings and capital gain payout of 10 respectively 14 Mio. CHF happened. The NAV after payments is 490 Mio. and the NAV at the end of 2006 is 515 Mio. CHF. The adjustment factor then is

$$f = \frac{490 + 10 + 14}{490} = 1.04898 \ .$$

This gives the performance for 2006:

$$P = \frac{515 * 1.04898}{500} - 1 = 8.045\%$$

There are several reasons why is important to measure the **performance of a fund** correctly. First, one wants to select the best fund. This leads to the question to which extent are the returns of a fund predictable. If the returns follow a random walk they are not predictable at all and the selection of the fund is a pure gamble. Second, to the funds stock to what they promised? Third, a correctly measured performance allows to check whether the fund manager added value.

- **Legal environment.** The legal environment is crucial for the development of the fund industry. About 3/4 of all cross-border funds in Europe are sold in Luxembourg. For private equity funds 2/3 has the State of Delaware as domicile and for Hedge Funds 1/3 are on the Caymans, 1/4 in Delaware. Luxembourg for example offers favorable framework conditions for holdings / holding companies, investment funds and asset-management companies. These companies are partially or completely tax-exempt; typically, profits can be distributed tax-free. As of Q3 2011, 48 % of the funds had their domicile in the US, followed by 9 % in Luxembourg, Brazil, France and Australia which each around 6 % . Luxembourg and Ireland are also specialized in setting up UCITS compliant structures, see below.

- **Distribution channel.** Funds can be privately offered or publicly distributed. The latter one is ment to sell the funds to all type of clients. The first one is used for expert investors. Regulation to protect the investor is heavy for publicly distributed funds.

Figure 3.26 shows the Swiss collective investment schemes. It shows that there are funds which are regulated and governed by the Swiss law (the CISA or KAG in German). Structured products are regulated by the CISA bus not governed. This status is likely to be changed in the near future. On a deeper level the open-end and closed-end funds are classified into four different subclasses. *'In the case of open-ended CIS, the CISA now draws a distinction between retail funds open to the public and funds for qualified*

The major benefits of funds for investors are:

- Diversification: Investors can buy a broadly diversified portfolio with small investment amounts.
- Investor protection: A large part of the fund industry faces strong regulation requirements.
- Access to assets: Funds allow investors investment into asset classes which would be impossible on a stand alone basis.
- Transparency: Retail funds in Europe make the investment process, the performance, the investment portfolio and the fees transparent.
- Default remoteness: Fund capital is treated as segregated capital.
- Investment strategy: The investor can choose between active and passive investment, can have access to short strategies, to rule based strategies, etc.

Contrary to structured products the payoff of a fund at a future date is not promised in exact mathematical terms, i.e. investors in funds belief that the fund management will generate a positive return due to its skills and information access.

investors. Public advertising is defined as any advertising aimed at the public. If advertising is aimed at qualified investors, it is not deemed to be public. A simplified approvals process is envisaged for both, but with different terms and conditions.' Source: SFA, 2012 . In Switzerland qualified investors are institutional investors such as banks, insurance companies, pension funds and private clients with wealth of at least CHF 2 million (not including real estate and pension amounts). A SICAV (Societe d'Investissement a Capital Variable) is an open-ended collective investment scheme. They are investment companies with a variable share capital that at all times equals the net asset value of the fund. SICAV are the most frequent legal form used. SICAVs are cross-border marketed in the EU under the UCITS directive, see below for UCITS. SICAF is the analogue to SICAV for a closed-ended funds. SICAR ia s Societe d'Investissement en Capital a Risque, i.e. a kind of venture fund and SIF, Specialized Investment Fund, are specialized vehicle funds.

The investment process for actively managed funds is a dynamic process which has a **tactical** and a **strategic** component. The objective of the tactical and strategic allocation is to obtain return above a **benchmark return**, which is often a passive portfolio. We use the optimization framework to show under which conditions the tactical and strategic components arise from a optimization problem, i.e. that they are not pure ad hoc concepts, see 3.42. The strategic allocation, by definition, is based on unconditional information whereas the tactical one is based on conditional information. The first one is based on historical returns and the second one on prediction models which use information today to forecast asset returns. This information processing view

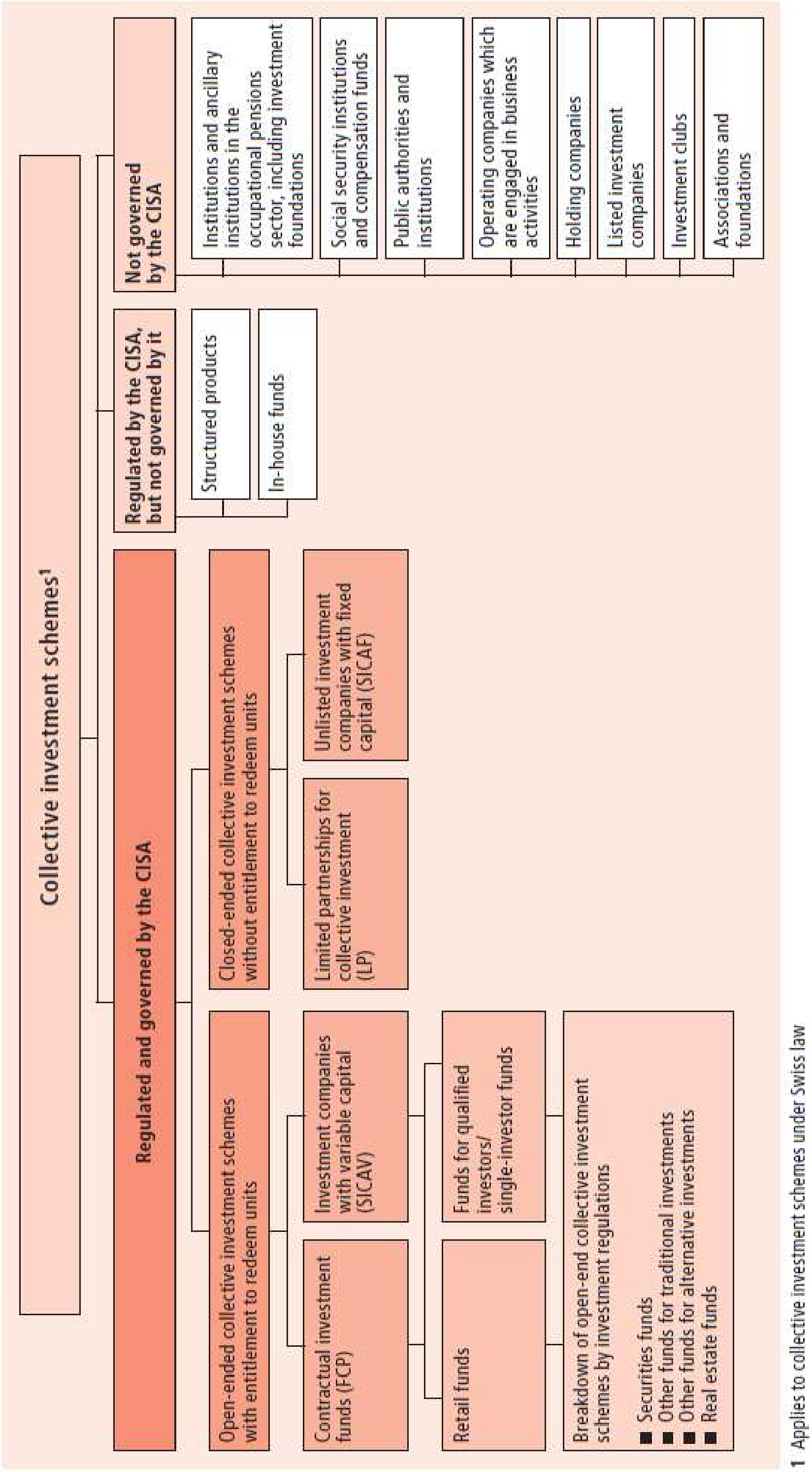


Figure 3.26: Swiss collective investment schemes.*Source: SFA, 2012*

leads for the strategic part to slowly varying portfolio weights (strategic) compared to the dynamic weights for the tactical allocation.

Figure 3.27 shows the returns of different asset classes which form the tactical asset allocation. The highest performing asset class of each year changes from time to time, there is no single asset that would always be the winner and the relative performance of asset classes vary from year to year, giving rise to return potential from asset class selection. Are these source exploitable? If markets are efficient, they are not. But linkages between these assets classes bases on economic facts make them exploitable. The table shows that a tactical asset allocation should be indeed dynamic and that not only the selection of the long asset classes but the short positions is key. The difference between the passive benchmark index and the return of the asset allocation is called the **total tracking error**.

3.6.2 European Fund Industry

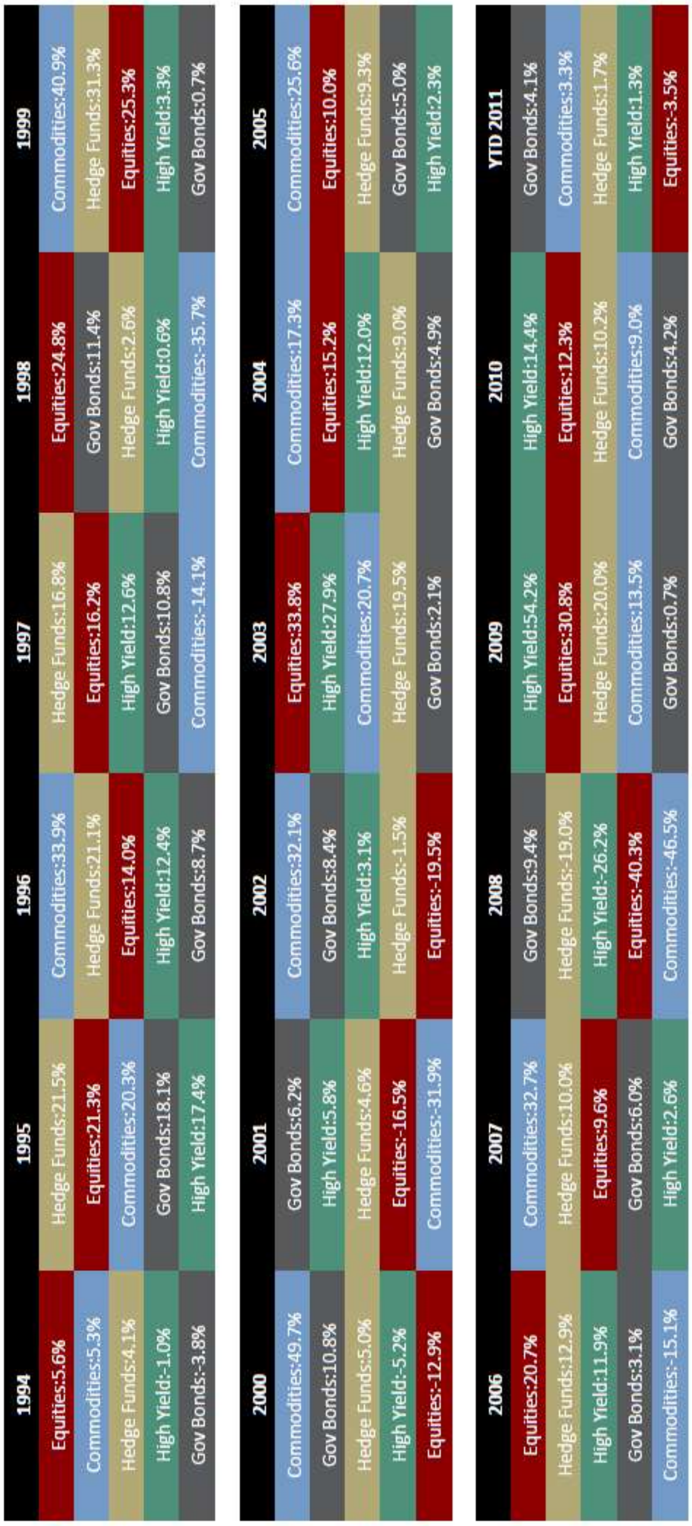
To obtain an impression about the size of the European fund industry, see Table 6.1.

Total Net Assets EU 27	7.6 Trillion EUR
Total No. of Funds EU 27	$\sim$ 50'000
Average Fund Size	1.01 Billion EUR
No. 1 Luxembourg	2 Trillion EUR
No. 2 France	1.4 Trillion EUR
No. 3 Germany	1.1 Trillion EUR
No.4 Ireland	0.97 Trillion EUR
No.5 UK	0.74 Trillion EUR
No.6 Switzerland	0.26 Trillion EUR

Table 3.5: Some figures of the European fund industry.

Luxembourg created an environment to attract different kind of funds by providing different kinds of vehicle to pool their investments. Luxembourg offers funds a choice of both regulated and non-regulated structures. If fund promoters establish a regulated fund in Luxembourg two options are available: (i) Undertaking for Collective Investment (UCI) which itself distinguishes between UCIs whose securities are distributed to the public and securities which are reserved for institutional investors and (ii) Société d'Investissement en Capital à Risque (SICAR) Both schemes fall under the prudential supervision of the Luxembourg financial sector regulator, the Commission de surveillance du secteur financier. The most common legal form of UCI is SICAV. A major reason for the Luxembourg fund industry is **taxation**. Luxembourg investment funds are not subject to income / capital gains taxes in Luxembourg. For withholding tax, taxation of capital gains also favorable taxation holds.

Figure 3.28 shows the different actors. The *investor* is the public with exception for US citizens. He receives dividends, redeems his shares and votes at SICAV meetings.



1994	1995	1996	1997	1998	1999
Equities:5.6%	Hedge Funds:21.5%	Commodities:33.9%	Hedge Funds:16.8%	Equities:24.8%	Commodities:40.9%
Commodities:5.3%	Equities:21.3%	Hedge Funds:21.1%	Equities:16.2%	Gov Bonds:11.4%	Hedge Funds:31.3%
Hedge Funds:4.1%	Commodities:20.3%	Equities:14.0%	High Yield:12.6%	Hedge Funds:2.6%	Equities:25.3%
High Yield:-1.0%	Gov Bonds:18.1%	High Yield:12.4%	Gov Bonds:10.8%	High Yield:0.6%	High Yield:3.3%
Gov Bonds:-3.8%	High Yield:17.4%	Gov Bonds:8.7%	Commodities:-14.1%	Commodities:-35.7%	Gov Bonds:0.7%

2000	2001	2002	2003	2004	2005
Commodities:49.7%	Gov Bonds:6.2%	Commodities:32.1%	Equities:33.8%	Commodities:17.3%	Commodities:25.6%
Gov Bonds:10.8%	High Yield:5.8%	Gov Bonds:8.4%	High Yield:27.9%	Equities:15.2%	Equities:10.0%
Hedge Funds:5.0%	Hedge Funds:4.6%	High Yield:3.1%	Commodities:20.7%	High Yield:12.0%	Hedge Funds:9.3%
High Yield:-5.2%	Equities:-16.5%	Hedge Funds:-1.5%	Hedge Funds:19.5%	Hedge Funds:9.0%	Gov Bonds:5.0%
Equities:-12.9%	Commodities:-31.9%	Equities:-19.5%	Gov Bonds:2.1%	Gov Bonds:4.9%	High Yield:2.3%

2006	2007	2008	2009	2010	YTD 2011
Equities:20.7%	Commodities:32.7%	Gov Bonds:9.4%	High Yield:54.2%	High Yield:14.4%	Gov Bonds:4.1%
Hedge Funds:12.9%	Hedge Funds:10.0%	Hedge Funds:-19.0%	Equities:30.8%	Equities:12.3%	Commodities:3.3%
High Yield:11.9%	Equities:9.6%	High Yield:-26.2%	Hedge Funds:20.0%	Hedge Funds:10.2%	Hedge Funds:1.7%
Gov Bonds:3.1%	Gov Bonds:6.0%	Equities:-40.3%	Commodities:13.5%	Commodities:9.0%	High Yield:1.3%
Commodities:-15.1%	High Yield:2.6%	Commodities:-46.5%	Gov Bonds:0.7%	Gov Bonds:4.2%	Equities:-3.5%

Figure 3.27: Returns of different components of a tactical asset allocation. *Source: Goldman Sachs, 2012.*

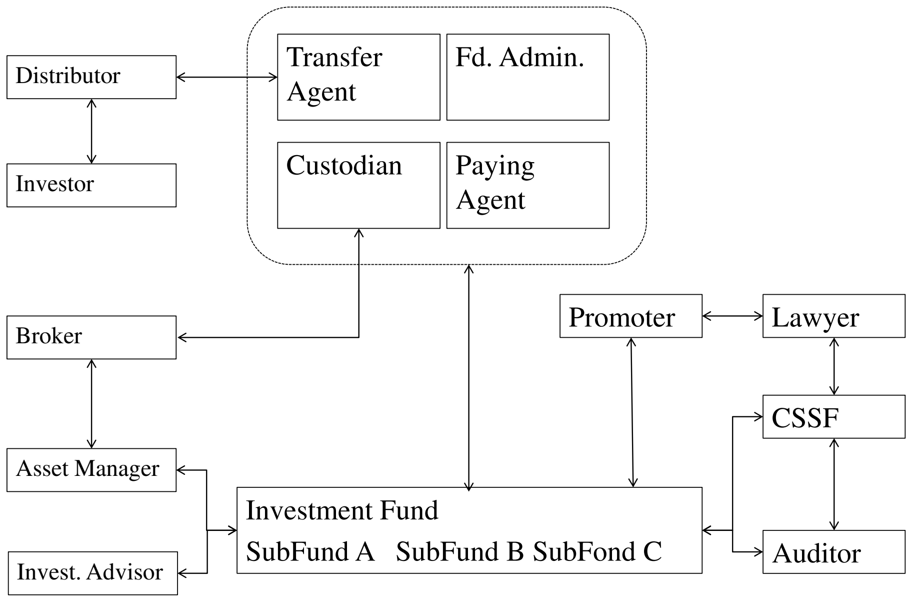


Figure 3.28: Actors in the fund industry. *Source: PricewaterhouseCoopers, 2011*

The *Board of Director* does the daily management of the fund, is responsible for risk management, reporting and transaction authorization. The *distributor* is an independent network who buys and sells shares/units and ensures marketing support. The *investment advisor* provides as a professional information and recommendations for the asset management, and creates buy and sell orders. The *promotor* creates as a professional of the wealth management sector in a bank the legal and administrative structure, distributes the fund, manages and controls the structures and provides services to the funds. The *transfer agent* is either the promoter or a third party. He executes authorized transactions, keeps and updates the register of the share units, issues certification of the issued shares. The *paying agent* is the promoter or a third party. He pays the dividends. *Finesti* collects, manages and disseminates financial information related to the Luxembourg-registered funds. *CSSF* is the Commission de Surveillance du Secteur Financier. They agree on the **prospectus**, the identity of the managers, the management and the custodian. A prospectus describes a the type and the risks of a security. Before a security is publicly offered, the prospectus has to be published in the EU. Different units in the banking firm - tax, legal, compliance, accounting, investor relations, risk, structuring - collaborate with external lawyers to write the prospectus. Errors, missing information or ambiguous information in the prospectus define a liability claim. Therefore the prospectus is an extensive document which is of little us for investment advice. The *Auditor* attests the accounting data of the annual report and provides the supervisory body with information and weaknesses.

The world wide largest *asset managers* are BlackRock (US) with 3.3 Trillion USD assets, State Street Global (US) with 1.9 Trillion USD, Allianz Group (D) with 1.9 Trillion USD, Fidelity Investments (US) with 1.7 Trillion USD and number 13 is UBS (CH) with 0.87 Trillion USD. *Custodians* (about 70 actors) safe keep the assets, perform daily managements of assets and monitor/control the operations. Leading with 467 billion of assets is JP Morgan, followed by State Street 265 billion, Dexia, Paribas, UBS. *Administrative Agent* is the promotor, a third party which has to be located in Luxembourg. They calculate the NAV, prepare the reports, contact the tax and regulatory authorities and do the accounting. Leading is JP Morgan, followed by Paribas, Dexia, State Street, UBS. The *Domiciliation Agent* is the promoter or a third party. He provides an offices, prepares meetings, distributes documents.

UCITS is an acronym for Undertaking in Collective Investment in Transferable Securities. This is the main European framework ruling investment funds. It organizes the distribution of investment funds with the purpose to protect investors. There were 4 framework initiatives; the last one UCITS IV initiated in 2009 (UCITS I 1985).

UCITS I

- The objectives were a harmonized legal framework to facilitate **cross-border** offering of investment funds to retail investors and to develop an integrated European market. The second objective was the definition of levels for investor protection (investment limits, capital organization, disclosure requirements, asset safe keeping, fund oversight). UCI funds do not have a passport property.
- The product passport facilitates the circulation of investment funds: Funds could be publicly distributed on a cross-border basis without satisfying the local regulations. This initiative turned out to be not fully satisfactory: Procedures for cross-border marketing take too long, are too costly and subject to too much supervisory interference. These issues were reconsidered in UCITS IV.

UCITS III

- In UCITS III the scope of investment funds was enlarged (derivatives) which facilitated the growth of the European fund industry.
- Further a management company is needed for portfolio and risk management, the administration and marketing.

UCITS IV

- UCITS IV has three main objectives: Reduce the administration burden by a notification procedure, increase the investor protection by a Key Investor Information (KID) and increase market efficiency.
- Before UCITS IV to distribute a fund abroad faced a waiting period of 2 months. With this new scheme the waiting period is reduced to 10 days. This facilitates the

product passport. With UCITS IV also management company passport is possible, i.e. a bank in France issues a foreign fund and can now distribute this foreign fund in Italy.

- The KID describes risks, returns, fee, historical performance figures in a standardized way. KID replaces the simplified prospectus, i.e. a less comprehensive form of the prospectus which could be used for some more standard or traditional types of funds.

- The master-feeder structures are ment to increase market efficiency. One or more investment vehicles pool their portfolio within another vehicle, i.e. there are several smaller feeder funds and one master to which they contribute. The asset pooling works as follows:

  - Feeder funds invest at least 85 percent of their assets in a master fonds.
  - To avoid cascade effects, the master feeder is not allowed to be a feeder fund neither to invest in feeder fonds.
  - The investment strategy of the feeder fund needs approval of a domestic authority.
  - Feeder and master funds are in a contractual relationship.
  - The structure has to be made transparent in the prospectus.

  The master-feeder structure allows asset managers to capture the efficiencies of larger pools of assets (economics of scale) although fashioning investment funds to separate market niches.

UCITS dominate global fund distribution in more than 50 local markets (Europe, Asia, Middle East, Latin America). This kind of global fund distribution is the preferred business model in terms of economies of scales and competitiveness. In summary, UCITS funds are (i) open-ended (investors can redeem shares/units), (ii) collective investment of capital raised from the public, (iii) diversified, (iv) investment in securities or other liquid financial assets, (v) retail oriented and (vi) product passport in 27 EU countries. UCITS funds make 89 percent and SIF 11 percent in terms of net assets.

3.7 Modern Portfolio Theory

3.7.1 Markowitz Model

The model of Markowitz (1952) derives optimal rule for the wealth allocation across risky assets when investors care only about the mean and variance of a portfolio's return. In this one period model the resulting portfolio weights are **not robust**, i.e. small variations in the estimated first and second moment (mean and variance) lead to large differences in the optimal asset allocation. More precisely often optimal portfolios depend in a spiky kink-type way on the parameters of the model such that a small variation in the parameters leads to a large change in the optimal portfolio. This instability is related to

estimation risk as follows. To obtain the optimal portfolio one needs estimates for the expected return and asset covariance in the Markowitz model. It follows that the sample estimates have a poor performance out-of-sample, i.e. there is uncertainty about the 'true' estimated values. Furthermore small changes in the expected return estimate lead to large changes in the portfolio weights. Data in the literature state figures that a one percent change of the estimate leads to a 200 percent change in the weight of an asset in the optimal portfolio. Similar effects hold for the covariance estimate.

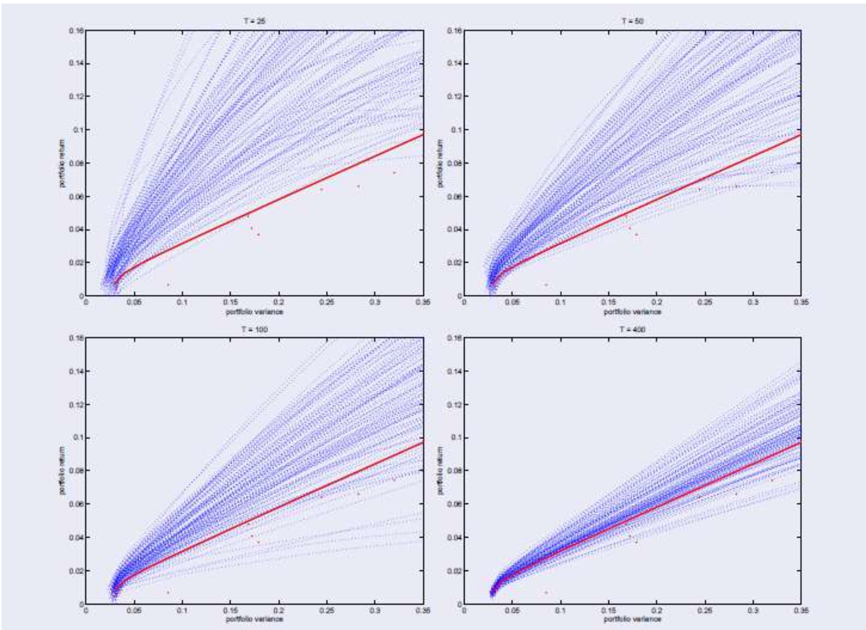


Figure 3.29: Estimation risk for mean-variance optimization. The figures shows that for increasing length of the time series the variability of estimation risk shrinks. The bold solid lines follow by inserting the sample estimates into the optimal policies. The other lines represent efficient frontiers if small parameter expected return and covariances are used.*Source: M. Leippold, Resampling and Robust Portfolio Optimization, 2010, University of Zurich.*

To understand this figure, we discuss the mean-variance (MV) approach of Markowitz. The approach is based on optimizing the trade-off between risk and return in a single period using the variance as the risk measure. A main advantage of this approach is its assumption that risk and return are related to each other and that a portfolio choice should be optimal with respect to these two objectives.

Consider two dates; 0 the trading date and T the performance measurement date . Asset prices at time 0 are known and exogenously given. Asset prices at time T are in general unknown at time 0. Either all asset prices at time T are random or one asset is singled out whose future price at time T is already known at time 0. We write B for this risk less asset and the time 0 price is normalized to $B(0) = 1$. The price at T of the risk less asset is assumed always larger than 1 in order to imply a positive interest rate. We further introduce N risky investment opportunities with known prices $S_j(0) \geq 0, j = 1, \ldots, N$, at time 0. Their prices $S_j(T), j = 1, \ldots, N$ at time T are non-negative random variables.

A **normalized portfolio** at time 0 is defined by (we omit the time index since the we consider a one period model)

$$\phi = (\phi_0, \ldots \phi_N) \ , \ \phi_0(0) = \frac{\psi_0}{V^\psi} \ \phi_k = \frac{\psi_k S_k}{V^\psi} \ , \ k = 1, \ldots, N \ . \tag{3.18}$$

with ψ_0 the amount of CHF invested in the saving account, ψ_j the number of units of the risky security j held in the period and the value process $V^\psi(t) = \psi_0 B(t) + \sum_{j=1}^N \psi_j S_j(t)$.

We write $R_j(0,T) =: R_j$ for the linear return of a security j and the return of a portfolio is the weighted sum

$$R^\phi = \phi_0 R + \sum_{j=1}^N \phi_j R_j \ . \tag{3.19}$$

where $\sum_{i=0}^N \phi_j = 1$. Using the scalar product notation the mean return and the covariance of a portfolio ϕ of N risky assets read:

$$E[R^\phi] = \langle \mu, \phi \rangle \ , \ var(R^\phi) = \langle \phi, V\phi \rangle \tag{3.20}$$

where μ is the vector of expected returns and V is the returns covariance matrix:

$$\begin{aligned} \mu_j &= E[R_j] \\ V_{ij} &= \begin{cases} E[R_i^2] - E^2[R_i] & \text{if } i = j \\ E[R_i R_j] - E[R_i]E[R_j] & \textit{else} \ . \end{cases} \end{aligned} \tag{3.21}$$

For two risky assets the expressions read:

$$\begin{aligned} \mu_\phi &= E[R^\phi] = \phi_1 \mu_1 + \phi_2 \mu_2 \\ \sigma_\phi^2 &= \mathrm{var}(R^\phi) = \phi_1^2 \mathrm{var}(R_1) + \phi_2^2 \mathrm{var}(R_2) + 2\phi_1\phi_2 \mathrm{cov}(R_1, R_2) \ . \end{aligned} \tag{3.22}$$

It is convenient to work with a'normalized covariance matrix', i.e. with the **correlation matrix** ρ, instead of the standard one V:

$$\rho = (\rho_{jk})_{1 \leq j,k \leq N} := \left(\frac{V_{jk}}{\sqrt{V_{jj} V_{kk}}} \right)_{1 \leq j,k \leq N} = \left(\frac{\mathrm{cov}(R_j, R_k)}{\sqrt{\mathrm{var}(R_j)\mathrm{var}(R_k)}} \right)_{1 \leq j,k \leq N} \ .$$

How does the standard deviations and correlations determine the covariance matrix? The covariance V can be written as the product $V = D'CD$. D is a diagonal matrix of the variances and C the matrix of correlations. Consider two assets with $\sigma_1 = 20\%$ and $\sigma_2 = 15\%$. Correlation is assumed to be 80 percent. Then,

$$D = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.15 \end{pmatrix} , \; C = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$$

which implies:

$$V = \begin{pmatrix} 0.04 & 0.024 \\ 0.024 & 0.00225 \end{pmatrix} .$$

Which properties should a reasonable covariance matrix V possess? The definition so far only implies that the matrix V is symmetric. The covariance matrix V needs to satisfy for all portfolio strategies ϕ the inequality

$$\langle \phi, V\phi \rangle \geq 0 \; , \; \forall \phi \; , \tag{3.23}$$

i.e. V has to be **positive semi definite** else variances become negative. See Appendix 7.7 for the proof.

Consider a portfolio of two equally weighted stocks with returns 2 and 3 percent and a covariance matrix

$$V = \begin{pmatrix} 0.01 & 0.005 \\ 0.005 & 0.015 \end{pmatrix} .$$

V is symmetric and the eigenvalues $0.150178, 0.00982166$ are positive, i.e. V is a covariance matrix. The expected returns are

$$\mu_\phi = E[R^\phi] = 0.5\,(0.02 + 0.03) = 0.025 \; .$$

The variance of the portfolio returns is

$$\sigma_\phi^2 = (0.5)^2 \times 0.01 + (0.5)^2 \times 0.015 + 2 \times (0.5)^2 \times 0.005 = 0.00875 \; ,$$

which implies a portfolio volatility σ_ϕ of 9.35 percent.

We always assume that the covariance matrix is strictly positive definite, i.e. we have for all portfolios ϕ

$$\langle \phi, V\phi \rangle > 0 \; . \tag{3.24}$$

A basic theorem of linear algebra states that a strictly positive definite matrix is invertible: V^{-1} exists. This will be used in the Markowitz model to solve for the desired optimal portfolios in closed form. What does this discussion about positivity of the covariance matrix implies? If the matrix is constructed from historical data one has to ensure and thus test for positivity. If this does not hold, the matrix under scrutiny is *not* a covariance matrix. If we consider a market with 10 securities, how can we test for the strict positivity of a given matrix? A theorem of linear algebra is states that a real symmetric matrix is strictly positive definite if and only if all eigenvalues are distinct and strictly positive. Consider the following matrices:

$$V_1 = \begin{pmatrix} 1 & 0.2 & 0.3 \\ 0.2 & 1 & 0.1 \\ 0.3 & 0.1 & 1 \end{pmatrix} \ , \ V_2 = \begin{pmatrix} 1 & -0.2 & -0.3 \\ -0.2 & 1 & -0.1 \\ -0.3 & -0.1 & 1 \end{pmatrix}$$

The eigenvalues of the two matrices are both strictly positive, i.e. strict positivity of a matrix has nothing to do with the signs of the matrix entries.

How can a times series of asset prices lead to a non positive definite covariance matrix, i.e. how is it possible that market data lead to portfolio of assets with negative variance? One reason are data which are not synchronous. For example different geographical regions have different holidays. Hence, correlation between two assets are matched at different dates. Another reason is the length of the time series. At its extreme consider a stock that has just been issued.

Example: We consider three securities, i.e. Standard& Poor's index 500, US Government Bonds and a US Small Cap Index. On a monthly basis the expected returns are over the period 1990-2004:

$$\mu = \begin{pmatrix} 0.0101 \\ 0.00435 \\ 0.0137 \end{pmatrix} = \begin{pmatrix} \text{Standard\& Poor} \\ \text{US Gov. Bonds} \\ \text{Small Cap Index} \end{pmatrix}$$

and the covariance matrix V is

$$V = \begin{pmatrix} 0.0032 & 0.0002 & 0.0042 \\ 0.0002 & 0.0004 & 0.0001 \\ 0.0042 & 0.0001 & 0.0076 \end{pmatrix} \ .$$

It follows that the bonds possess the smallest expected return but also the smallest risk. The eigenvalues of this covariance matrix are

$$0.01019 \ , \ 0.00073 \ , \ 0.00045 \ .$$

Therefore, V is a strictly positive definite, symmetric matrix. The largest eigenvalue belongs to the Small Cap Index. This eigenvalue is much larger than the two others. This reflects that the volatility or riskiness of this stock market source dominates the two other ones.

We derive systematically the efficient frontier, i.e. we state and solve the **Markowitz model**. We assume:

1. There are N risky assets and no risk free asset. Prices of all assets are exogenous given.

2. There is a single time period. Hence, any intertemporal behavior of the investors can not be modelled.

3. There are no transaction costs. This assumption can be easily relaxed nowadays since a Markowitz model with transaction costs can be numerically solved.

4. Markets are liquid for all assets. This assumption, which also essentially simplifies the analysis, is much more demanding to remove than the absence of transaction costs restrictions.

5. Assets are infinitely divisible. Without this assumption, we would have to rely on integer programming in the sequel.

6. Full investment holds i.e.
$$\langle e, \phi \rangle = 1$$
with $e = (1, \ldots, 1)' \in \mathbf{R}^n$.

7. Portfolios are selected according to the mean-variance criterion.

We note that short selling is allowed. So far we consider the optimization program

$$\begin{array}{lll} \min\limits_{\phi \in \mathbf{R}^n} & \frac{1}{2}\langle \phi, V\phi \rangle & (\mathbf{M}) \\ s.t. & \langle e, \phi \rangle = 1 & \\ & \langle \mu, \phi \rangle = r\ . & \end{array}$$

This model is a quadratic optimization problem and the feasibility set, i.e. the two constraints, is convex since it is the intersection of two hyperplanes. The factor $\frac{1}{2}$ is chosen for notational convenience and the solution of the program depend on the parameter r, the **excess return**. Does the problem has a solution and is it unique? With the additional assumptions, general non-linear optimization theory gives a positive answer to both questions:

1. The covariance matrix is strictly positive definite.

2. The vectors e, μ are linearly independent.

3. All first and second moments of the random variables exist.

The Markowitz model $\mathbf{M}$ consists of the quadratic optimization program, the assumption 1.-7. and the technical assumptions 1.-3.

The unique solution is given next.

Proposition 3.7.1. *If the above assumptions hold, the solution of the model* $\mathbf{M}$ *is*

$$\phi^* = r\phi_0^* - \phi_1^* \qquad (3.25)$$

with

$$\begin{aligned}
\phi_0^* &= \frac{1}{\Delta}\left(\langle e, V^{-1}e\rangle V^{-1}\mu - \langle e, V^{-1}\mu\rangle V^{-1}e\right)\\
\phi_1^* &= \frac{1}{\Delta}\left(\langle e, V^{-1}\mu\rangle V^{-1}\mu - \langle \mu, V^{-1}\mu\rangle V^{-1}e\right)\\
\Delta &= |W^{-1}e|^2|W^{-1}\mu|^2 - (\langle W^{-1}e, W^{-1}\mu\rangle)^2\\
V &:= WW' \ , \ |x| := \sqrt{\langle x, x\rangle} \qquad (3.26)
\end{aligned}$$

where W is an $N \times N$ matrix.

We note that ϕ_0^* and ϕ_1^* are independent of r. See the Appendix 7.7 for the proof. We define the following expressions:

$$a = \langle \mu, V^{-1}\mu\rangle \ , \ b = \langle e, V^{-1}e\rangle \ , \ c = \langle e, V^{-1}\mu\rangle \ .$$

Proposition 3.7.2. *In the model* $\mathbf{M}$*, the following bounds hold:*

$$b \in \left[\frac{N}{\lambda_{max}}, \frac{N}{\lambda_{min}}\right] \ , \ c \in \left[\frac{|\mu|^2}{\lambda_{max}}, \frac{|\mu|^2}{\lambda_{min}}\right] \ , \ |a| \leq \frac{\sqrt{N}|\mu|}{\lambda_{min}} \qquad (3.27)$$

with λ_{max} (λ_{min}) the maximum (minimum) eigenvalue of V and a, b, c are all positive.

See the Appendix 7.7 for the proof. We note that the eigenvalues of the covariance matrix represents the volatility of the risk sources of the different assets - and not the volatility of the assets itself. The eigenvalues are the entries of the diagonal matrix Λ, i.e. the matrix $\Lambda = D'VD$ which exists according to the Spectral Theorem of Linear Algebra. Therefore, one eigenvalue λ_1 dominates all other eigenvalues, the corresponding risk can used as a proxy model where only this risk factor is considered. If all risk factors are similar, i.e. $\lambda_j \sim \lambda_i$ holds, then the maximum and minimum eigenvalue are of the same size too. Therefore, the intervals for b, c shrink and also the admissible values for a become smaller. Since these three parameters describe the hyperbola of the optimal portfolio variance, see (3.7.1) below, similar importance of risk sources shift the hyperbola in $(\sigma(r), r)$-space, see Figure ??, as follow:

- The hyperbola is shifted to the right and upwards since increasing a means a shift to the right (to obtain the same expected return more risk is required) and c shifts the efficient frontier upwards (with the same risk level higher expected returns are possible). Risk increases in this case and in particular global minimum variance is higher. A decreasing value of b reduces the curvature of the frontier, i.e. it becomes flatter which means that an increase of one unit of risk leads to a weaker increase of expected return.

- If the maximum eigenvalue increases, i.e. one risk source is becomes more and more dominant, then b, c both increase. The admissible intervals for both parameters increase. This shifts the frontier upwards to the right.

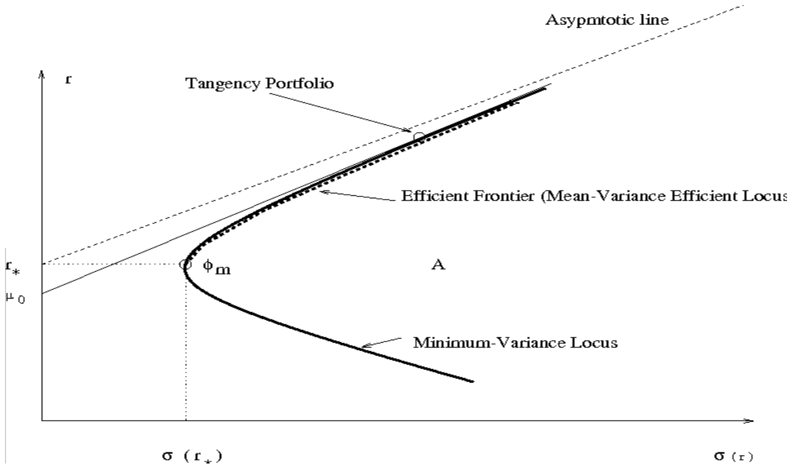


Figure 3.30: Mean-variance illustrations in the $(\sigma(r), r)$-space. The asymptotic line is given by $r = \sigma + r_* \sqrt{\langle \phi_0^*, V\phi_0^* \rangle}$. The tangency portfolio and the tangent follow from the Markowitz problem with a risk less asset, see discussion below.

We calculate the variance for the optimal portfolio ϕ^* of Proposition 3.7.1:

$$\sigma^2(r) \quad = \quad \langle \phi^*, V, \phi^* \rangle = \frac{1}{\Delta} \left(r^2 b - 2rc + a \right) \ . \tag{3.28}$$

The locus of this set in the $(\sigma(r), r)$-space are hyperbolas (see Figure ??). Proposition 3.7.1 implies that the function

$$\sigma(r) = \sqrt{\langle \phi^*, V\phi^* \rangle}$$

provides the minimum standard deviation and variance for any given mean r. Since V is positive definite, the quadratic form under the root is convex and positive. Since the root is strictly increasing on the positive real numbers, the function $\sigma(r)$ is convex. Therefore, the function has a unique minimum r_* which is given by the solution of

$$\sigma'(r) = 0 \ \Rightarrow r_* = \frac{c}{b} \ .$$

Example: We consider three stocks with the expected return and covariance matrix:

$$\mu = (0.2, 0.3, 0.4) \ , \ V = \begin{pmatrix} 0.1 & 0.08 & 0.09 \\ 0.08 & 0.15 & 0.07 \\ 0.09 & 0.07 & 0.25 \end{pmatrix} \ .$$

This is indeed a covariance matrix since all eigenvalues are strictly positive $(0.3429, 0.1224, 0.03459)$. The inverse matrix V^{-1} is:

$$V^{-1} = \begin{pmatrix} 22.4363 & -9.42877 & -5.43703 \\ -9.42877 & 11.6311 & 0.137646 \\ -5.43703 & 0.137646 & 5.91879 \end{pmatrix} \ .$$

This leads to the values $a = 0.922918, b = 10.5299, c = 2.46387$ and $\Delta = ab - c^2 = 3.6476$. The optimal portfolio components then are:

$$\phi_0^* = \begin{pmatrix} -6.60377 \\ 3.20755 \\ 3.39623 \end{pmatrix} \ , \ \phi_1^* = \begin{pmatrix} -2.26415 \\ 0.528302 \\ 0.735849 \end{pmatrix} \ .$$

The efficient frontier is:

$$\sigma(r) = 0.27415(0.92291 - 4.92773r + 10.52993r^2) \ .$$

Suppose that the investor wants a return of $r = 30\%$. Sufficient is then to invest in stock number 2. But we can obtain the same return by inserting $r = 0.3$ in the optimal investment strategies:

$$\phi = \begin{pmatrix} 0.283019 \\ 0.433962 \\ 0.283019 \end{pmatrix} \ , \ \sigma(0.3) = 0.107 \ .$$

The investor is optimally long the first stock by an amount of 28.3 percent of his wealth and similar for the other assets. The risk of this optimal portfolio is 10.7 percent which is much less than the 15 percent from an investment in asset 2 only.

Using the results of Proposition 3.7.1 we calculate the **global minimum variance portfolio** $\phi_m^* = \phi_m^*(r_*)$ and the **global minimum variance** for this strategy :

$$\phi_m^*(r_*) \quad = \quad \frac{1}{b} V^{-1} e \ , \ \sigma(r_*)^2 = \frac{1}{b} \ . \tag{3.29}$$

The global minimum variance is **independent** of the return properties of the assets.

The following remarks are immediate. First, for all portfolios on the efficient frontier there exists no other portfolio with the same mean and a lower standard deviation. In other words the portfolios on the efficient frontier are not *dominated* by any other portfolio in A. Second, for each inefficient portfolio exists an efficient portfolio with the same variance but a higher expected rate of return. Third, any minimum variance portfolio is an efficient portfolio if:

$$r \geq \frac{c}{b} = r^* \ .$$

Consider the case of two securities with $\mu_1 = 1, \mu_2 = 0.9$ and

$$V_{11} = 0.1, V_{22} = 0.15, V_{21} = V_{12} = -0.1 \ .$$

The expected return r is equal to $r = 0.96$. The two assets are negatively correlated and if we neglect the correlation structure, asset 2 does not seems to be attractive since it possesses a lower expected return and a larger risk measured by the variance. However, the negative correlation will induce that it can be advantageous to invest in the second asset too. First we consider the strategies $\phi_1 = (1,0), \phi_2 = (\frac{1}{2}, \frac{1}{2})$. We get

$$var(R^{\phi_1}) = 0.1 \ , \ E(R^{\phi_1}) = 1 \ , \ var(R^{\phi_2}) = 0.0.0125 \ , \ E(R^{\phi_1}) = 0.95 \ .$$

Although ϕ_1 satisfies the expected return condition $r = 0.96$, the risk is much larger than for the strategy ϕ_2 which in turn does not satisfy the expected return condition. As a final strategy we consider ϕ_3 which is obtained by solving the optimization problem without imposing any restrictions on the expected return. It follows that $\phi_3 = (\frac{5}{9}, \frac{4}{9})$ is the searched portfolio and

$$var(R^{\phi_3}) = 0.011 \ , \ E(R^{\phi_3}) = 0.955 \ .$$

Hence the risk is minimal but the expected return is smaller than $r = 0.96$. Finally, if we solve the full Markowitz problem, the optimal portfolio reads $\phi^* = (0.6, 0.4)$ and we get

$$var(R^{\phi^*}) = 0.012 \ , \ E(R^{\phi^*}) = 0.96 \ .$$

Therefore, 40 percent has to be invested in the not very attractive asset. This is the **Markowitz phenomenon**: To reduce the variance as much as possible, a combination of negatively correlated assets should be chosen. For this portfolio, compared to the naive one ϕ_1, the variance is reduced drastically and the expected return is still acceptable.

Figure 3.31 shows the efficient frontier for SMI stocks in the year 2000.

Two frontiers are shown, once with short selling constraints and once without such constraints. It follows that for a given expected return level higher risk level is needed in cause with such constraints. This shows the fact that each constraint not only reduces potential losses but also reduces the return potential.

So far, we considered the return restrictions $\langle \mu, \phi \rangle = r$. What happen if we instead consider the constraint $\langle \mu, \phi \rangle \geq r$? Suppose, that $\langle \mu, \phi \rangle > r$ holds. By the slackness condition in the KKT theorem, the corresponding multiplier for this constraint is zero. This simplifies the optimization problem and we get for the optimal policy

$$\hat{\phi}^* = \frac{V^{-1}\mu}{c} \ . \tag{3.30}$$

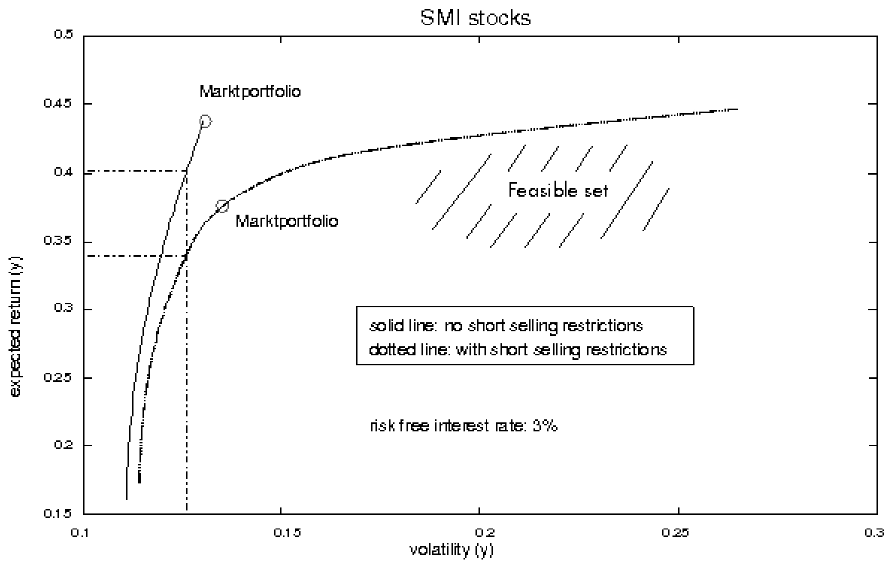


Figure 3.31: The efficient frontiers for two Markowitz-type models are shown: The Markowitz model without and with short selling restrictions. Data: SMI, daily, year 2000.

With this portfolio, an arbitrary optimal portfolio $\phi^*(r)$ reads

$$\phi^*(r) = \nu(r)\phi^*_m + (1 - \nu(r))\hat{\phi}^* \ . \tag{3.31}$$

The function $\nu(r)$ can be determined by the representation of $\phi^*(r)$ in Proposition 3.7.1. We already found two decompositions of the optimal minimum variance portfolio in two portfolios. In fact, any minimum variance portfolio $\phi^*(r)$ can *always* be written as a combination of two linearly independent minimum variance portfolios. Since for any portfolio, there exists another one such that a weighted combination is equal to the optimal portfolio, the decomposition is called *the mutual fund theorem* in the literature.

Proposition 3.7.3 (Mutual Fund Theorem)**.** *Any minimum variance portfolio can be written as a combination of the global minimum variance portfolio and the portfolio $\hat{\phi}^*$ given in (3.30). Furthermore, any minimum variance portfolio is a combination of any two distinct minimum variance portfolios.*

The second statement means that we can replace the global minimum variance portfolio and $\hat{\phi}^*$ by any other combination of minimum variance portfolios. See the Appendix 7.7 for the proof. The practical implication are as follows. Suppose that an investor wishes to invest optimally according to the mean-variance criterion. A possibility is to buy the assets such as prescribed by the portfolio $\phi^*(r)$. But if the number of assets is large this might be impossible or impractical to achieve. The decomposition property allows the investor to search for two funds with the characteristic of the two portfolios in the decomposition (3.31) and which are simpler to buy. Then the investor only has to invest in these two funds to buy the optimal portfolio. The next proposition characterizes the efficient frontier in terms of the expected returns, variances and covariances of the returns.

Proposition 3.7.4. *Consider the Markowitz model* $\mathbf{M}$ *and* $r \geq \frac{c}{b}$*. A portfolio* ϕ *is efficient if and only if a positive affine relation between the covariance of the return of each asset* R_i *with the portfolio* R^ϕ *and the expected return exists. Formally,*

$$cov(R_i, R^\phi) = f_1^\phi E[R_i] + f_2^\phi \ , \ f_1^\phi \geq 0 \ , \ i = 1, \ldots, N \ . \tag{3.32}$$

The proof is given in Appendix 7.7.

To interpret the condition (3.32) we use the following result in the Markowitz model $\mathbf{M}$:

$$\frac{\partial \sigma^2(r)}{\partial \mu_k} = cov(R_k, R^{\phi^*}) \tag{3.33}$$

holds with ϕ^* a minimum variance portfolio.

The impact of one unit more return in asset k on the optimal variance equals the covariance of asset k with the minimum variance portfolio. If the asset k is positively correlated with the portfolio, a unit more return of this asset increases the variance and the contrary holds, if the correlation is negative. Hence, to reduce the variance as much as possible, a combination of negatively correlated assets should be chosen, i.e. the

Markowitz phenomenon. The condition (3.32) can be interpreted as follows: A portfolio is mean-variance efficient if and only if the marginal contribution of each security to the portfolio risk is an positive, affine function of the expected return.

We rewrite condition (3.32) in a form which is used in practice.

Proposition 3.7.5. *Consider the Markowitz model* $\mathbf{M}$ *and* $r \geq \frac{c}{b}$*. A portfolio* ϕ *is efficient if and only if - with expectation of the global minimum variance portfolio - there are uncorrelated, arbitrary portfolios* $\bar{\phi}$*, such that*

$$E[R_i^{\phi}] = E[R_i^{\bar{\phi}}] + cov(R_i, R^{\phi}) \frac{E[R^{\phi}] - E[R^{\bar{\phi}}]}{\sigma^2(R^{\phi})} \ , \ i = 1, \ldots, N \tag{3.34}$$

with $E[R^{\phi}] - E[R^{\bar{\phi}}] > 0$.

The slope of the affine relationship between expected return and the covariance is $\frac{E[R^{\phi}]-E[R^{\bar{\phi}}]}{\sigma^2(R^{\phi})}$. One defines $\beta_i = \frac{cov(R_i, R^{\phi})}{\sigma^2(R^{\phi})}$ and the vector of Betas

$$\beta = (\frac{cov(R_N, R^{\phi})}{\sigma^2(R^{\phi})}, \ldots, \frac{cov(R_N, R^{\phi})}{\sigma^2(R^{\phi})})'$$

is a measure of the covariance of the assets with the efficient portfolio normalized by the variance of the efficient portfolio.

So far all assets in the Markowitz problem were assumed to be risky. We assume that there exists a **risk less** asset B_t with return μ_0 apart of the N risky assets. All other properties of the economy and of the financial market are left unchanged. The optimization problem is stated in the Appendix 7.7. It follows, that in $\sigma(r), r)$-space the locus of the minimum-variance frontier is a straight line, see Figure ??. The next proposition characterizes minimum variance portfolios and specifies a natural representation of minimum variance portfolios.

Proposition 3.7.6.

1. *All minimum variance portfolio are combinations of any two linearly independent minimum variance portfolios.*

2. *The minimum variance portfolio in the model* $\mathcal{M}_R$ *with zero investment in the risk less asset (*$\phi_0 = 0$*) is given by*

$$\phi^T = \frac{1}{c - \mu_0 b} V^{-1}\mu - \frac{\mu_0}{c - \mu_0 b} V^{-1} e \ . \tag{3.35}$$

   This portfolio is called the tangency or market portfolio. The tangency portfolio is also an efficient portfolio in the model $\mathcal{M}$.

3. *The efficient portfolios of the model* $\mathcal{M}_R$ *on the branch* $r = \mu_0 + \sigma\sqrt{\Delta_R}$ *are tangent to the efficient frontier of the model* $\mathcal{M}$.

The proof is omitted.

We note without further comments:

- Assume the models $\mathcal{M}$ and $\mathcal{M}_R$ given and $\mu_0 \neq r^*$. Then the tangency portfolio is element of the efficient frontier if and only if $\mu_0 < r^*$.
- The tangency portfolio is element of the inefficient part of the minimum variance locus if and only if $\mu_0 > r^*$.
- If $\mu_0 = r^*$, the tangency portfolios does not exists for any finite risk and return. The portfolios are then given by the intersection of the asymptotic portfolios and the efficient set at infinity.

We know that any minimum variance portfolio is a linear combination of two distinct minimum variance portfolios. Contrary to the risky-asset only case, there is a natural choice of the mutual funds: The risk less asset and the fund with no risk less asset, i.e. the tangency portfolio ϕ^T. A necessary and sufficient condition for portfolio efficiency - analogous to the risky asset only case - can be given.

Proposition 3.7.7. *Consider the model $\mathcal{M}_R$. For a portfolio ϕ being an efficient portfolio necessary and sufficient is*

$$E[R_i^\phi] = \mu_0 + cov(R_i, R^\phi)\frac{E[R^\phi] - \mu_0]}{\sigma^2(R^\phi)} \ , \ i = 1, \ldots, N \tag{3.36}$$

with $E[R^\phi] - \mu_0 > 0$.

The proof is omitted.

3.7.2 CAPM

So far, portfolio theory considered investors' asset demand given asset returns. How investors' asset demand determines the relation between assets' risk and return in a market equilibrium, i.e. when demand equals supply? We ask for a model which prices the assets. We consider the Capital Asset Pricing Model (CAPM) and the main question is to price the assets in equilibrium is: What is the expected return of the assets? That for we recall that the risk less asset and the tangency portfolio ϕ^T are the natural investments to hold and that an asset's risk premium is proportional to its systematic risk:

$$E[R_i^\phi] - \mu_0 = \beta_{i,T}[E[R^T] - \mu_0]$$

with $E[R^T]$ the expected return of the tangency portfolio, μ_0 the risk free return and $\beta_{i,T}$ the Beta between i and the tangency portfolio. If we can **identify** the tangency portfolio, we obtain the asset pricing model. In the CAPM we also need the market portfolio, ie the portfolio of all traded securities. The market capitalization M_i of asset i equals the price multiplied by the outstanding title. M denotes total market capitalization and ϕ_i^M the weight of title i in the market portfolio. The CAPM assumes:

- Investors have the same opinions about the distribution of returns.

- Investors have the same fixed investment horizon.
- Investors hold efficient frontier portfolios.
- There is a risk-free securities in zero net supply.
- Demand for the securities is in equilibrium equal to the supply (market clearing).
- All information is available at the same time to all investors.
- Markets are perfect: There are no trading costs, lending and borrowing is not limited under the risk free rate.

Proposition 3.7.8. *In the above model we have in the equilibrium:*

- *Each investor is investing in two classes: The risk less asset and the tangency portfolio.*
- *The tangency portfolio is the market portfolio.*
- *All investors hold the same portfolio of risky securities.*
- *For each title i we have:*
$$\mu_i - \mu_0 = \beta_i(\mu^T - \mu_0) \ , \tag{3.37}$$
with $\mu_i = E[R_i^\phi]$.

If the risk-free rate and the Betas are known the expected returns follow. The proof is simple. Consider I investors each endowed with wealth e^j. Investor n is invested with e_0^n in the risk less asset, the rest is invested in the tangency portfolio. The sum of this risk less positions over all investors is zero due to the zero net supply assumption. The second equilibrium condition states that wealth invested in the tangency portfolio of all investors is equal to the market portfolio weighted by the market capitalization, i.e.

$$\sum_n (e^n - e_0^n)\phi^T = M\Phi^M \ .$$

Since the second sum is zero, we have $\sum_n e^n = M$. This implies $\phi^T = \Phi^M$: Hence total wealth is invested in the risky assets and the tangency portfolio equals the market portfolio.

Consider an economy with three risky securities A, B, C and three investors, see Table 3.6. The tangency portfolio is $\phi^T = (0.2619, 0.3095, 0.4286)$, investors endowment is $250, 300, 500$ and the portfolios of the investors are given in the table. Since in equilibrium the value of each security is equal to its Market value. Thus, the last line in the table summarizes the market capitalization of 1050. The market portfolio then reads.

$$\phi^M = (275/1050, 325/1050, 450/1050) = (0.2619, 0.3095, 0.4286) = \phi^T \ .$$

Investor	risk less	A	B	C
1	50	50	50	100
2	-150	150	200	100
3	100	75	75	250
sum	0	275	325	450

Table 3.6: Portfolios of investors.

Relation (3.37) represents the expected return μ as a function of the Betas. The graph of this map is called the **Security Market Line (SML)**. The relation (3.37) allows to decompose the asset return in three components:

$$\mu_i - \mu_0 = \alpha_i + \beta_i(\mu^T - \mu_0) + \epsilon_i \ ,$$

with the noise term ϵ a standard normal one which is not correlated to the market portfolio. The Alpha should be zero according to the CAPM.

Some assumptions of the CAPM are criticized. The Markowitz model assumes quadratic utility functions (mean and variance). It can be shown that the result of Markowitz holds for general utility function but then the asset returns need be normally distributed. But some asset returns are not normally distributed - they show fat tail behavior for example and the assumption of quadratic preferences is a very specific choice to hold true for all investors. Several effects, such as the size or value effect cannot be explained by CAPM. This led the factor models initiated by Fama and French. Variance is often criticized to be a meaningful measure stick of risk since it punishes the upside and fails to measure drawdowns. We comment below on a drawdowns risk measure (Value-at-Risk). There is a lot of critique from a behavioral finance point of view. CAPM assumes that the probability beliefs of active and potential shareholders match the true distribution of returns. Behavioral finance studies the possibility that active investor's expectations are biased, causing market prices to be informationally inefficient. CAPM does not appear to adequately explain the variation in stock returns. Empirical studies show that low beta stocks may offer higher returns than the model would predict. A final remarks concerns the market portfolio. This portfolio should include all types of assets that are held by anyone as an investment - from liquid assets to illiquid ones as real estate or even more opaque one as art. Such a market portfolio is unobservable. One substitutes a stock index as a proxy for the true market portfolio. Unfortunately, Roll (1977) showed that this substitution can lead to false inferences as to the validity of the CAPM. The CAPM might not be empirically testable. This is referred to as Roll's critique.

3.7.3 Equivalent Formulation of Markowitz Model, Tactical and Strategic Asset Allocation

We always minimized the portfolio variance under return and other constraints. One could **equivalently** maximize the return under a risk constraint. A possible model is

$$\begin{aligned} \min_{\psi} \quad & \frac{1}{2}\langle \psi, V\psi \rangle - \delta \langle \mu, \psi \rangle \qquad , (\mathcal{M}_1) \\ s.t. \quad & \langle \psi, e \rangle = 1, \psi \in \mathbf{R}^N, \end{aligned} \tag{3.38}$$

with δ an exogenous parameter. The solution of this model is

$$\psi^* = V^{-1}(\delta\mu + \lambda e) \ , \ \lambda = \frac{1-\delta c}{b} \ . \tag{3.39}$$

The two portfolios ψ^* and ϕ^* (i.e. the solution of the original model) are equivalent, iff

$$\langle \psi^*, \mu \rangle = \langle \phi^*, \mu \rangle \ , \ \sigma^2_{\phi^*}(r) = \sigma^2_{\psi^*}(\delta) \ , \tag{3.40}$$

where in the risk measure the explicit dependence on the respective model parameters are shown. A short calculation shows that the variance condition is always satisfied iff

$$r = \frac{\delta\Delta + c}{b} \ . \tag{3.41}$$

To prove this, we calculate

$$\langle \psi^*, V\psi^* \rangle = \rho^2 \frac{\Delta}{b} + \frac{1}{b}$$

and compare it with

$$\langle \phi^*, V\phi^* \rangle = \frac{1}{\Delta}(r^2 b - 2rc + a) \ .$$

The linear relation between r and δ in (3.41) guarantees that the expected returns and the variances in both models agree. Therefore, for any choice of desired expected return r there exist a value δ such that the model $\mathcal{M}_1$ leads to the same outcome. This proves that the two models are equivalent.

Although the two models are mathematically equivalent they may *not* be equivalent from a behavioral point of view. Consider an investor which plans to find out its optimal one-period investment according to the mean-variance criterion. In the model $\mathcal{M}$ he needs then to fix the desired rate r, whereas in the model $\mathcal{M}_1$ he has to determine the trade-off between risk and return by selecting the parameter δ. The equivalence of the models designs an experiment for a laboratory to test peoples rationality in decision making under uncertainty. Another equivalent formulation is:

$$\max_{\phi} \left(\langle \mu, \phi \rangle - \frac{\gamma}{2} \langle \phi, V\phi \rangle \right)$$

with γ the investor's risk aversion.

We consider the transition from the Markowitz model to the Capital Asset Pricing Model (CAPM). This allows us to provide a rational foundation for the tactical and strategic asset allocation. We recall that the tactical asset allocation considers short-term investments. The difference between the return on the benchmark index and the return on the tactical asset allocation is the total tracking error. We start with the quadratic optimization:

$$\max_{\psi} \quad \langle \mu, \psi \rangle - \frac{\gamma}{2} \langle \psi, V\psi \rangle \; s.t. \; \langle \psi, e \rangle = 1 \; . \tag{3.42}$$

The solution of this problem (Mutual Fund Theorem) can be written as a combination of the global minimum variance portfolio $\phi_m^* = \frac{V^{-1}e}{b}$ and a second portfolio $\hat{\phi}^* = \frac{V^{-1}\mu}{c}$ That is,

$$\psi^* = (1 - \nu(r))\phi_m^* + \nu(r)\hat{\phi}^*$$

holds with the weight $\nu(r) = c/\rho$. Rearranging this is equivalent to

$$\psi^* = \phi_m^* + \frac{V^{-1}}{\gamma} (\mu - b/c \times e)$$

where e is the vector with 1 in each component. The equilibrium model CAPM delivers a vector of equilibrium returns $\tilde{\mu}$. By adding and subtracting this vector in the last expression and rearranging we get:

$$\psi^* = \phi_m^* + \frac{V^{-1}}{\gamma} \left(\tilde{\mu} - \tilde{b}/c \times e \right) + \frac{V^{-1}}{\gamma} \left(\mu - \tilde{\mu} - (b - \tilde{b})/c \times e \right)$$

where $\tilde{b} = \langle 1, V^{-1}\tilde{\mu} \rangle$. Rearranging further we get:

$$\psi^* = \phi_m^* + \phi_S^* + \phi_T^*$$

with

$$\phi_S^* = \frac{1}{\gamma} \left(\frac{V^{-1}(\tilde{\mu}e - e\tilde{\mu})V^{-1}}{c} \right) \times e \tag{3.43}$$

$$\phi_T^* = \frac{1}{\gamma} \left(\frac{V^{-1}((\mu - \tilde{\mu})e - e(\mu - \tilde{\mu}))V^{-1}}{c} \right) \times e \tag{3.44}$$

We end up with three components, a global minimum variance component, a strategic component, and a tactical component. Note that $\tilde{\mu}e - e\tilde{\mu}$ is a matrix with zero in the diagonal and $\tilde{\mu}_j - \tilde{\mu}_i$ as entry in the cell ij - i.e. it is the equilibrium excess return of asset i versus asset j. Similarly, the cell in the matrix for the tactical allocation represents the deviation of expected excess return of asset i versus asset j from equilibrium excess return in the cell ij.

If the asset returns in equilibrium are all the same the strategy component is zero and hence irrelevant. In this case only tactical bets and the global minimum variance portfolio matter. The size of both the tactical and the strategic components also depend on the risk aversion degree and the covariance matrix. The tactical strategy is a strategy of pairwise bets which are weighted by the inverses of the covariance matrices, i.e. bets between less volatile assets are larger in size than between more risky assets.

3.7.4 Mean-Value-at-Risk Portfolios, Risk Measures

A critique of the mean-variance criterion for optimal portfolio selection often concerns the variance. First, the variance is a symmetric measure of risk. But why penalize the upside in the portfolio selection. Second, the variance is often not seen as a true measure of risk. That is, the measure fails to detect the states which reflect a economic distress. To improve the situation, one extensively considered **mean-value-at-risk (VaR)** portfolio optimization. The VaR VaR_α at a level $1-\alpha$ is defined for the portfolio return R^ϕ by:

$$P(R^\phi \le -\mathrm{VaR}_\alpha) \ge \alpha \ .$$

It is the minimum amount an investor can loose (in dollars) with a confidence interval of $1-\alpha$. The bigger the VaR, the more risky the portfolio is. The definition has implicit a time horizon assumption: For trading portfolios the portfolios are left unchanged for 10 days, i.e. the VaR is calculated on a ten day basis. This reflects the idea that the risk is calculated by assuming that the portfolio positions are unchanged for 2 trading weeks. For credit risk, the time horizon is often one year due to the lower liquidity to change say mortgage positions. A new efficient frontier can be calculated using the mean and the VaR. For normal (more general elliptical) distributions the efficient frontier and the optimal portfolios can be calculated explicitly although the VaR risk measure is defined only implicitly.

Proposition 3.7.9. *The portfolio returns R^ϕ are normally distributed with mean μ and volatility σ. The VaR then reads*

$$-VaR_\alpha = \sigma k_\alpha + \mu \ .$$

The critical figure k depends on the confidence level $1-\alpha$ and is tabulated.

The proof is given in Appendix 7.7. Given the simplicity of the VaR under normality it is no longer a big surprise that one can solve the optimization problem explicitly. In fact, the optimization is again the maximization of a quadratic function under linear constraints. Why? The portfolio variance $\sigma^2 = \langle \phi, V\phi \rangle$ is inserted in the VaR-expression. That for, one takes the square of the VaR constraint in order to switch from the volatility to the variance. This produces the quadratic term. Doing a lot of algebra one derives the optimal mean-VaR portfolio and finds that the mean-VaR efficient frontier in a mean-standard deviation framework is given by a mean-variance efficient frontier: Under the assumption of multivariate distribution for assets' returns, one can find for every mean-variance efficient portfolio, which differs from the global minimum variance portfolio a confidence level (if it exists) such that this portfolio corresponds to the global minimum Value-at-Risk portfolio, see Figure 3.32.

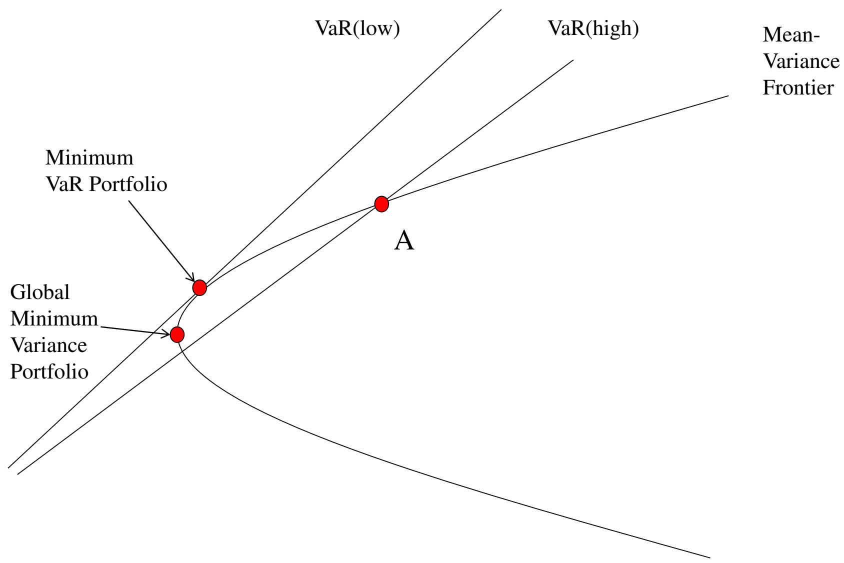


Figure 3.32: Mean-variance and mean-VaR portfolios. The Figure shows (Point A) that for every mean-variance efficient portfolio there is a confidence level which leads to the same portfolio. The figure indicates that with a decreasing confidence level the VaR is moving to the left, i.e. the lower risky portfolio follow in the equivalent mean-variance setup. It also follows that there is a lower bound on the confidence level such that for lower levels no equivalent mean-variance optimal portfolio exists. Besides the confidence level also the VaR amount level matter.

As in the mean-standard deviation framework, one can derive the minimum VaR portfolio the tangency portfolio which is the e portfolio that maximizes the ratio mean/VaR. It gives the portfolio with maximal return per unity VaR.

As an applications of VaR consider an Euro investor. The investor has the following portfolio:

Position	**Type**	**Market Price**	Currency
1	10 Equity Funds Shares DAX	1'000	Euro
2	5 Equity Funds Shares DJ	5'000	USD
31	200 Novartis Stocks	50	CHF
4	10 US Treasury, 10y, Zero Coupon Bonds	800	USD

Table 3.7: Initial value of the investor's portfolio.

There are three equity risk sources (DAX, DJ, Novartis), two FX risks USDEUR (actual 1.05) and CHFEUR (actual 0.8) and interest rate risk for the bond, i.e. 6 risk factors. We calculate the **weekly VaR on a 95% level**. To calculate the VaR we need the variance and covariance information, see Figure 3.33, the calculation of the exposure in Euro and the allocation of the EUR exposure to the risk factors. The portfolio variance $\sigma_p^2 = X'CX$ is the product of the vector X of the EUR exposure allocated to the risk factors times the covariance matrix C (the entry ij in this matrix is $\sigma_i \sigma_j \rho_{ij}$). Calculating these matrix products gives $\sigma_p^2 = 160'804'032$. This is the value on an annual basis. To obtain the result on a weekly basis, we obtain for the variance

$$\sigma_w = \sqrt{160'804'032/52} = 1'758 \ .$$

The critical value on the 95% level is $k_{95\%} = 1.644853$. This implies the 1w EUR VaR of:

$$\text{VaR} = 1'758 \times 1.644853 = 2'892 \text{ EUR} \ .$$

If we consider the VaR expression $-\text{VaR}_\alpha = \sigma k_\alpha \sqrt{T} = \sqrt{X'CX} k_\alpha \sqrt{T}$, where we neglect the drift and insert the time-scaling, we get the VaR contribution:

$$\frac{\partial \text{VaR}}{\partial X} = k_\alpha \sqrt{T} \frac{CX}{\sqrt{X'CX}} = k_\alpha^2 T^2 \frac{CX}{\text{VaR}} \ .$$

This implies the contribution rule:

$$\text{VaR} = \sum_j X_j \frac{\partial \text{VaR}}{\partial X_j} = k_\alpha \sqrt{T} \sum_j X_j \frac{(CX)_j}{\sqrt{X'CX}} = \sum_j \text{VaR}_j \ . \tag{3.45}$$

If we apply this in the example we get the VaR contributions for the risk factors. The contribution of the US Treasury bond is negative, i.e. due to its negative correlations to the other factors the VaR is reduced by 6 percent. The largest VaR contribution is from the DAX risk factor with 31 percent, although the exposure is only 10.5 percent. The

Market Data	**Standard Deviation**	**Correlations** DAX	DJ	Novartis	USDEUR	CHFEUR	US 10y Treasury
DAX	30%	1	0.5	0.6	0.4	0.55	-0.2
DJ	20%	0.50	1	0.55	0.77	0.66	-0.4
Novartis	25%	0.60	0.55	1	0.23	0.72	-0.22
USDEUR	15%	0.40	0.77	0.23	1	0.73	-0.49
CHFEUR	5%	0.55	0.66	0.72	0.73	1	-0.21
US 10y Treasury	10%	-0.20	-0.4	-0.22	-0.49	-0.21	1

First Calculations Position	Type	Market Price	Euro Exposure	**Allocation of the EUR Exposure to the risk factors** DAX	DJ	Novartis	USDEUR	CHFEUR	US 10y Treasury
1	10 Equity Funds Shares DAX	1'000	10'000	10'000					
2	5 Equity Funds Shares DJ	5'000	26'250		26'250		26'250		
3	200 Novartis Stocks	50	8'000			8'000		8'000	
4	10 US Treasury, 10y, Zero Coupon Bonds	800	8'400				8'400		8'400
			Sum (Vector X)	10'000	26'250	8'000	34'650	8'000	8'400

Variances & Covariances C

	DAX	DJ	Novartis	USDEUR	CHFEUR	US 10y Treasury
DAX	0.090000	0.030000	0.045000	0.018000	0.008250	-0.006000
DJ	0.030000	0.040000	0.027500	0.023100	0.006600	-0.008000
Novartis	0.045000	0.027500	0.062500	0.008625	0.009000	-0.005500
USDEUR	0.018000	0.023100	0.008625	0.022500	0.005475	-0.007350
CHFEUR	0.008250	0.006600	0.009000	0.005475	0.002500	-0.001050
US 10y Treasury	-0.006000	-0.008000	-0.005500	-0.007350	-0.001050	0.010000

Figure 3.33: Data and calculations for the VaR example.

contribution of USDEUR is 19 percent to VaR whereas the factor exposure is the largest one of 36 percent.

We conclude with some remarks on **general risk measures**. This field of research starts with a number of properties a risk measure should have and then they prove whether a function, the risk measure, exists which satisfies the properties, whether the function is unique and important for practitioners if and how the risk measure can be explicitly constructed. In the last decade the focus was on **coherent** and **convex** risk measures. We discuss verbally the former one which was initiated by the work of Artzner et al. (1999).

The model starts with the following assumptions. First, the risk of holding no assets is zero. This is a normalization assumption. If a portfolio takes better values than another one in (almost all) states then the better portfolio risk should be less than the risk of the worse portfolio. Hence, the risk measure should be monotone. The risk of a portfolio is not larger than the risk of individual components. This is the diversification principle. Next, if one double the portfolio value then one doubles portfolio risk. Finally, if one adds risk less cash to a portfolio, the difference between the portfolio with cash minus the portfolio itself equals cash. A risk measure which satisfies these assumptions is called a coherent risk measure. The authors proved the existence of such a measure. It follows that VaR is only coherent if returns are normally distributed. More precisely VaR fails to honor diversification for general return distributions. An often used measure which is coherent is expected shortfall. This measure is based on the general not coherent VaR-measure. Expected shortfall measure the risk of a portfolio, given that the loss exceeds VaR. I.e. it is the expected loss value given that the loss exceed VaR. If the loss distribution under consideration is heavy tailed expected shortfall will be substantially larger than VaR since there is probability mass far away from VaR level. Contrary if a loss distribution is rapidly decreasing, expected shortfall and VaR are close together. The former case is that one considered in loan portfolios the other one in trading portfolios of liquid shares.

3.7.5 Resampling, Robust Optimization

The Markowitz model of last section is plagued by estimation risk. As market conditions change, portfolios need to be rebalanced to keep them optimal. In the mean-variance approach frequent rebalancing of the portfolio is necessary. The cause of the apparent need for rebalancing are not better investment opportunities, but statistical instability of some underlying estimates. This instability and ambiguity are a major problem in practical applications of mean-variance theory. By instability and ambiguity we mean that small changes in input often lead to large changes in the optimized portfolio.

In mean-variance analysis the mean μ and the covariance matrix V are estimated from time series. These estimates $\widehat{\mu}, \widehat{V}$ are different from the true population values: **estimation errors** result. As Scherer (2002) states: *'The optimizer tends to pick those* *assets with very attractive features (high return and low risk and/or correlation) and tends to short or deselect those with the worst features. These are exactly the cases where estimation error is likely to be highest, hence maximizing the impact of estimation error on portfolio weights. The quadratic programming optimization algorithm takes point estimates as inputs and treats them as if they were known with certainty (which they are not) will react to tiny differences in returns that are well within measurement error.'* This is the reason that mean-variance optimized portfolios suffer from instability and ambiguity. To reduce estimation risk, resampling is used. An algorithm is as follow:

1. Let $\widehat{x}_0 = (\widehat{\mu}_0, \widehat{V}_0)$ be the historical estimates.
2. Sample from multivariate normal distribution with the input $\widehat{x}_0 = (\widehat{\mu}_0, \widehat{V}_0)$.
3. Compute the new sample mean and covariance. Use them to create the efficient frontier.
4. Save portfolio weights that are on the new efficient frontier.
5. Repeat the steps 1. to 4. m times.
6. Average over saved portfolio weights to obtain final portfolio weights that lie on the resampled efficient frontier. Since averaging means smoothing, the saved average portfolio weights will be more smooth.

These sampled portfolio mean-variance pairs are below the theoretical efficient frontier which is the optimal one satisfying the mean-variance criterion. A plot shows that considerable variations follow unless the length of time series is very long (under iid returns). Figure 3.34 shows 8 asset weights in the mean-variance approach, the mean-variance frontier and the resampled frontiers.

The resampled efficient frontier is visually superior to the theoretical mean-variance efficient frontier. It is better diversified, i.e. allocation to all asset classes is present and transition between portfolios is continuous and smooth, no sharp changes follow. Resampled portfolio weights change in a smooth way as risk tolerance changes. Large differences may arise to the mean-variance weights. However, the efficient frontiers are very close.

Robust mean-variance optimization is another approach to control for estimation risk. The model is similar to the considered robust Merton problem of optimal consumption and investment. That is, one solves the problem

$$\max_{\phi} \left(\min_{\mu \in S_\mu, V \in S_V} \left(\langle \phi, \mu \rangle - \frac{\gamma}{2} \langle \phi, V\phi \rangle \right) \right) . \tag{3.46}$$

The set $S_.$ defines the sets of allowed model misspecification, one for the return and one for the covariance matrix. As in the Merton model, nature chooses the worst model for the investor. This leads to very conservative optimal investments, i.e. one is mostly invested in cash. To see this, assume that short positions are not allowed. Then the worst return vector in the set of all returns is that one with the lowest expected returns, i.e.

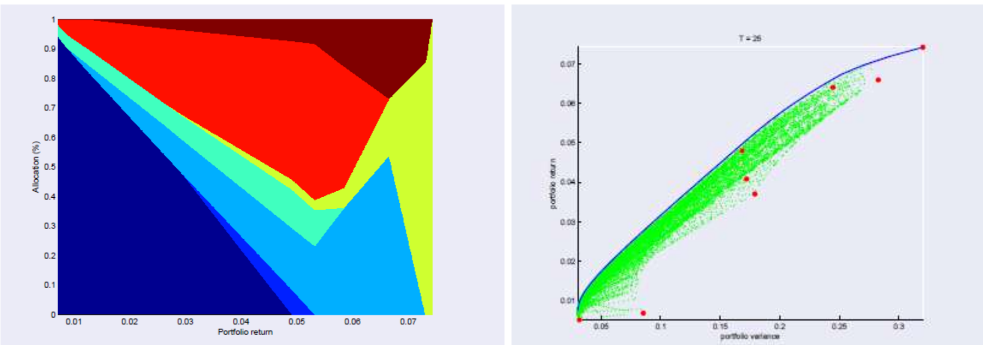


Figure 3.34: Left Panel: 8 asset weights in the mean-variance approach, the mean-variance frontier and the resampled frontiers. Right Panel: Resampled efficient frontier. The red dots represent the eight assets. *Source: M. Leippold, Resampling and Robust Portfolio Optimization, 2010, University of Zurich.*

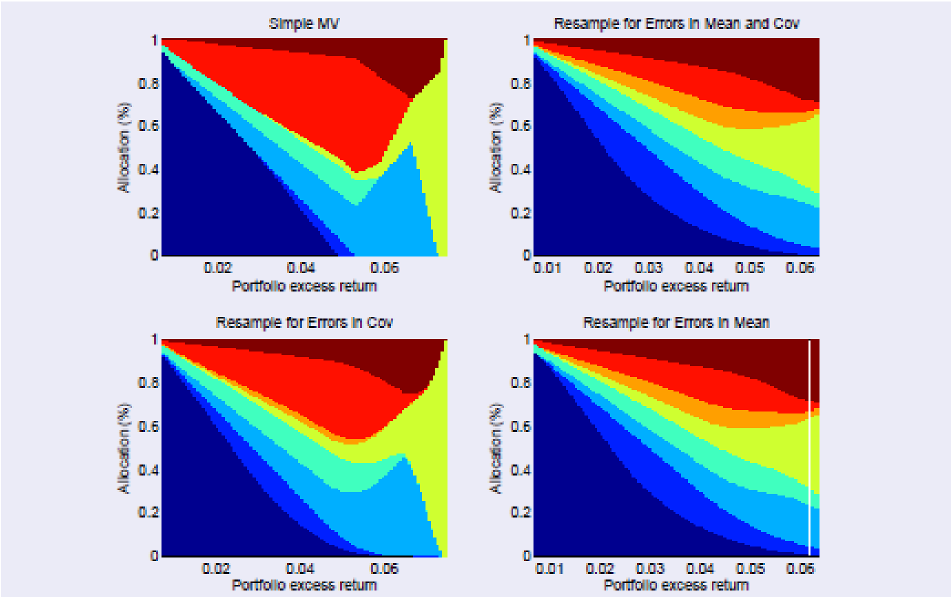


Figure 3.35: Smoothness of resampled portfolio weights. The weights are smoother in the resampled case. *Source: M. Leippold, Resampling and Robust Portfolio Optimization, 2010, University of Zurich.*

it is the smallest element in $\S_\mu$ measured by its norm. For the covariance the worst case means little diversification. But little diversification means that V is the matrix which is closest to the diagonal one or with the largest determinant. Therefore, the min-max problem reduces to a max-only problem where the return is replaced by the lowest one and the covariance by the largest one in the above sense. If we would allow for short positions, the argumentation breaks down since for example low returns become optimal for a short position. For any optimization with both risky assets and cash, we will end up with a 100 per cent cash holding as long as we look deep enough into the estimation error tail. This is overly pessimistic. This formulation is furthermore equivalent to very narrow Bayesian priors where investors would get the same result by putting a 100 per cent weight on the two above extremal return and covariance. Although there are several extension of the above robust decision problem, analysis shows that it adds not much to traditional mean-variance.

3.7.6 Optimal versus Naive Investment

The discussions of the last sections allow us to consider optimal versus naive investment. DeMiguel et al. (2009) ask under which conditions mean-variance optimal portfolio models can be expected to perform well even in the presence of estimation risk. They evaluate the out-of-sample performance of the sample-based mean-variance portfolio rule together with 13 model extensions which are designed to **reduce estimation error** relative to the performance of the naive portfolio 1/N diversification rule, i.e. the rule which allocates a fraction 1/N of wealth to each of the N assets at each rebalancing date. In other words, 1/N is a rule which does not care about statistical dependence. The **sample-based mean-variance portfolio** simply plugs in the estimated mean and variance in the optimal allocation rule. This portfolio strategy completely ignores the possibility of estimation error. The 13 model extensions include Bayesian approaches to estimation error, moment restrictions, portfolio constraints and optimal combinations of portfolios. The mixture portfolios are constructed by applying the idea of shrinkage directly to the portfolio weights. The out-of-sample performance relative to 1/N is compared across 7 different data sets of monthly returns using three performance criteria:

- the out-of-sample **Sharpe ratio**, i.e. the ratio of return per one unit of volatility.
- the **certainty-equivalent (CEQ)** return for the expected utility of a mean-variance investor, i.e. the risk free rate that an investor is willing to accept rather than adopting a particular risky portfolio strategy.
- the **turnover (trading volume)** for each portfolio strategy.

They show that of the 14 models evaluated (Markowitz plus 13 extensions), **none is consistently better** than the naive 1/N benchmark in terms of all three performance measures.

The authors apply the following method to evaluate the performance. The use a 'rolling-sample' approach. For a data set of asset return of length T, they choose an estimation window M of length 5 or 10 years. DeMiguel et al. (2009): *In each month t, starting from $t = M + 1$, we use the data in the previous M months to estimate the parameters needed to implement a particular strategy. These estimated parameters are then used to determine the relative portfolio weights in the portfolio of only-risky assets. We then use these weights to compute the return in month $t+1$. This process is continued by adding the return for the next period in the data set and dropping the earliest return, until the end of the data set is reached. The outcome of this rolling-window approach is a series of $T-M$ monthly out-of-sample returns generated by each of the portfolio strategies.'*

The authors find:

- Bayesian strategies do not seem to be very effective at dealing with estimation error.
- Constraints alone do not improve performance sufficiently.
- Combined portfolio constraints with some form of shrinkage of expected returns are usually much more effective in reducing the effect of estimation error.
- In only two cases are the CEQ returns from optimizing models statistically superior to the CEQ return from the 1/N model.
- No single strategy always dominates the 1/N strategy in terms of Sharpe ratio. The 1/N strategy has Sharpe ratios that are higher relative to the constrained policies, which have Sharpe ratios that are higher than those for the unconstrained policies.
- Only the passive strategy where the investor holds the market portfolio and does not trade at all is better than the 1/N strategy in terms of turnover.
- The simulation results show that for reasonable parameter values, the models of optimal portfolio choice reduce only moderately the critical length of the estimation window needed to outperform the 1/N policy. Long estimation windows are required before the sample-based mean-variance policy achieves a higher out-of-sample Sharpe ratio than the 1/N policy. For 10 risky assets, the Sharpe ratio of the sample-based mean-variance policy is higher than that of the 1/N policy only for the case of $M = 6000$ months; for 25 and 50 assets it does not achieve the same Sharpe ratio as the 1/N policy even for an estimation window length of 6000 months.

The intuition for their findings are: First, the vector of expected excess returns over the risk-free rate and the variance-covariance matrix of returns have to be estimated. This needs very long time series to estimate them precisely. Since the strategies are sensitives to the input parameters, small errors in the estimates yet leads to weights that are far from optimal. Hence, the error due to the naive 1/N can be smaller than the optimized model ones. Second, the 1/N rule performs well in the data sets because wealth is distributed across portfolios of stocks rather than individual stocks. Portfolios of stocks have lower idiosyncratic risk than individual stocks, i.e. the loss from naive as opposed to optimal diversification is much smaller when allocating wealth across portfolios. Optimal diversification policies dominate the 1/N rule only for very high levels of idiosyncratic volatility.

Tu and Zhou (2011) continue the analysis of the former authors. Using new data sets all tested sophisticated strategies not only under perform the 1/N, but also **have negative risk-adjusted returns**. Investors are worse off by using the rules than holding cash. The authors analyze whether a combination between the 1/N rule and with a sophisticated rule yields a better rule. The combination of two investment rules exactly obtains the average performance of two if one rule is good in some scenarios and bad in others and the other rule may do the opposite. Statistically, the combination can be interpreted as a shrinkage estimator with the 1/N rule as the target. A shrinkage defines a tradeoff between the bias and variance. The 1/N rule is biased, but with zero variance. A sophisticated rule is usually asymptotically unbiased, but has sizable variance in small samples. When the 1/N is combined with the sophisticated rule, an increase of the weight on the 1/N increases the bias, but decreases the variance. An optimal weight must make the combination better than either of the two rules on a stand alone basis. They show that the **Kan and Zhou (2007)** model coupled to the 1/N rule performs consistently well across models and data sets. It performs as well as or better than all other sophisticated rules on a consistent basis. It also outperforms substantially the 1/N across almost all models. Moreover, this combination rule never loses money (on a risk-adjusted basis) across models and data sets. Kan and Zhou (2007) propose a 'three-fund' portfolio rule, where the third fund minimizes 'estimation risk'. The intuition is that estimation risk cannot be diversified away by holding only a combination of the tangency portfolio and the risk-free asset. Hence with estimation risk an investor also benefits from holding some other risky-asset portfolio - a third fund. Kan and Zhou search for this optimal three-fund portfolio rule in the class of portfolios that can be expressed as a combination of the sample-based mean variance portfolio and the minimum-variance portfolio.

Chapter 4

Swaps and Financial Markets

Swaps are one of the most successful financial innovations in the last decades. Due to the importance of swaps and the moderate complexity we consider different aspects in swap innovation:

- Rationale.
- Pricing.
- Construction of the time-value of money curve.
- Standard documentation innovation (ISDA).

We also give a broad overview of capital markets, money markets and the classification of derivatives.

4.1 Introduction to Swaps

We introduce to the vast swap topic focussing on vanilla Interest Rate swaps (IRS)[1]. These bilateral contracts generalize Forward Rate Agreements (FRAs), see Figure 4.3. Most often fixed versus floating rate are exchanged. The reference rate for the floating leg is typically LIBOR or EURIBOR. The notional amount is not exchanged it serves only as a calculation figure. In USD or Euro the maturity range starts with 2 years up to 30 years. To enter such a contract an ISDA agreement is necessary and counter party risk limits are needed. The contractual size differs for different currencies. In Swiss Francs the minimum size is CHF 2 Mio. The day count conventions are different for the fixed and floating leg. For the fixed payments the day count convention of bond markets

[1]The notion vanilla is used for basic derivatives. Vanilla products on equity are call and put option. More complicated products are labelled exotic.

30/360 is used and act/360 is used for the floating leg, the money market basis.[2] The counter party paying the fixed rate is called the 'payer', the other one the 'receiver'. The payer (receiver) is by convention long (short) the swap.

Originally, IRS were introduced for **interest arbitrage** reasons. Consider two firms A and B. A has a high creditworthiness, B a low one. Both firms can borrow at a fixed or floating rate given in Table 4.1.

	A	B	Difference
Fixed	5%	6 %	1 %
Floating	LIBOR	LIBOR + 0.75 %	0.75 %

Table 4.1: Rates for firms A and B.

The two firms can both benefit if they enter into a IRS, since the differences in the fixed rate borrowing is different from the floating rate one. There is a mispricing of credit or liquidity risk or both of them in the market. Both parties can realize and divide the difference of 0.25% among them using an IRS as follows. To lock in the profit, each party borrows where they have an advantage. A borrows fixed and B floating. B agrees to pay A floating rate LIBOR plus 0.75 percent and A agrees to pay B fixed 5.9 percent. A gets floating rate funding at LIBOR minus 0.15 percent and B gets an advantage in fixed funding of 0.1 percent, Figure 4.1 illustrates the transaction.

The first swap was designed 1981 between the World Bank and IBM. The motivations of the two parties are shown in Figure 4.2:

IBM received DM and CHF from its funding program (we skip CHF). IBM used this money to finance project in the US, i.e. IBM exchanged DM debt in 1979 for dollar funds. That for IBM needed to change periodically USD in DM to serve the coupon payments. Since USD became stronger in that period compared to DM, IBM made currency gains. To realize these gains IBM needed to get ride of its DM-liabilities.

The World Bank borrowed in the capital markets and lent to developing countries for project finance. The costs of the loans were the same than the financing cost of the World Bank in the markets. US interest rates were at 17 percent in this period. The comparable rates in Germany and Switzerland were 12% and 8%. World Bank was raising funds in low interest rate currencies, such as the DM. Since the World Bank was constraint to borrow in Germany and Switzerland the bank needed to find another party which owed DM and which wanted to exchange them against USD - IBM.

An investment banker at Salomon Brothers realized that a **currency swap** would solve the problems of both parties. One party agrees to make interest payments in currency A at a fixed rate on a notional amount N and to pay that notional amount in

[2]'30/360' means that each month has 30 days and each has 360 days. The convention 'act' means that the actual calender dates are summed. The reset frequency is the frequency of floating payments.

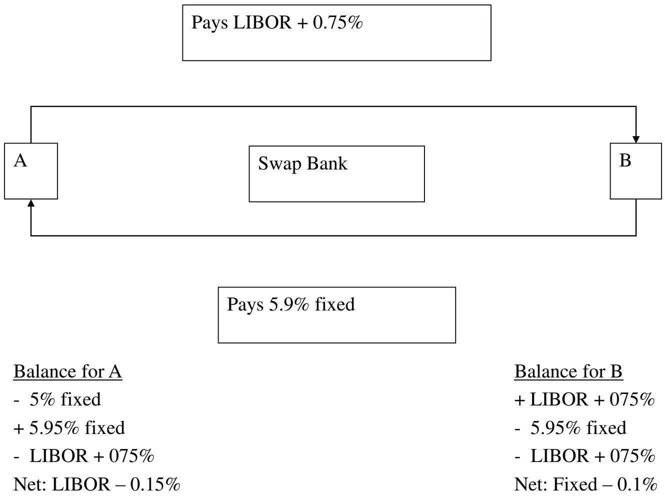


Figure 4.1: Interest rate arbitrage using IRS.

currency A at maturity of the swap to a counter party in exchange for receiving from it all interest rate payments in currency B on a notional amount N and receiving N in currency B at maturity. The notional amounts are chosen to have a value in a common currency at initiation, so that they do not have to be exchanged. Salomon Brothers managed the trade on a back-to-back basis where IBM could change their DM-liabilities into USD and the World Bank could buy DM at favorable rates: The World Bank lent IBM over notional amounts and coupons denominated in DM and received from IBM notional and coupons in USD. The success of this transaction was necessary to overcame skepticism against such a **product innovation**. The World Bank got out of its USD debt. The end result for IBM was as if it did not have DM but dollar debt.

After this trade in a second period in the 80's banks started to enter into own-name transactions. I.e. swap counter parties discussed directly with the bank as intermediary their desired risk and return profile. Entering between the two parties the bank faced twice counter party risk. One also stared to develop standardized documentation documents which allowed to process this taylor made transaction effectively, see ISDA agreements below. The third period was characterized by the beginning market making. Banks started to trade swaps with several counter parties. Market and counter party risk increased due to this wider activities - large investment in risk management followed. Market risk was often compensated with transactions in other markets. For example, due

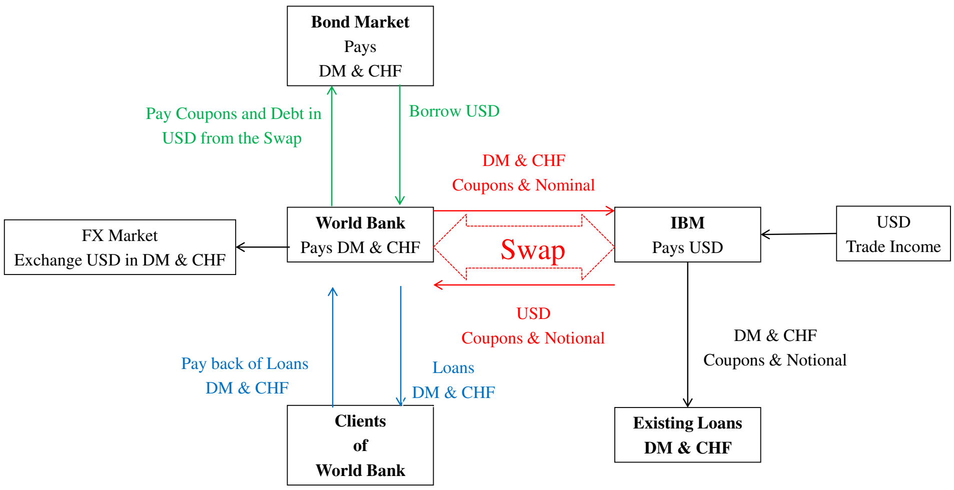


Figure 4.2: Swap between the World Bank and IBM.

to their liquidity government bonds were preferred.

We report some swap market figures in the Section Capital Markets.

4.2 Swap Pricing

To illustrate how the fixed rate is determined, designate the current date as time zero and the final maturity date of the swap as time T. The fixed rate at which a new swap with maturity T can be executed is the constant **par swap rate** $s_{0,T}(0)$. This rate **by definition sets the value of the swap at initiation to zero**, i.e.

$$\mathrm{PV}_{\mathrm{Swap}}(0, s_{0,T}(0)) = 0 \; .$$

The argument (0) refers to calendar time, the first subindex denotes the start date of the swap and T is the maturity date of the swap. Once a swap is executed, fixed payments of $s_{0,T}(0)$ are made annually. We assume that the difference between two consecutive fixed dates is equidistant. Floating payments are made quarterly. They are equal to act/360 times the 3m LIBOR rate at the beginning of the quarter. This is called **setting in advance and paying in arrears**. A floating rate note paying 3m LIBOR quarterly must be worth **par** at each quarterly LIBOR reset date, see below for a proof. Since the initial value of a swap is zero, the initial value of the fixed leg must also be worth

par. Figure 4.3 shows this using a graphical decomposition of a swap into par fixed and floating rate bonds. Solving for the swap rate in the above equality

The fixed and floating leg are both at par at initiation. We have

$$\mathrm{PV}_{\text{Swap, fixed}}(0, s_{0,T}(0)) = \mathrm{PV}_{\text{Swap, floating}}(0) \ .$$

$$s_{0,T}(0) = \frac{1 - p(0,T)}{A_{0,T}(0)} = \frac{\text{PV Floating}}{\text{Annuity}} \tag{4.1}$$

where $A_{0,T}(0) = \sum_{j=1}^{T} p(0,j)$ is the present value of an annuity with first payment six months after the start date and final payment at time T. This annuity is called the **level of the swap**. We prove (4.1). The fix part is simply the present value of a fix rate bond, i.e. $\sum_j s_{0,T}(0)p(0,T_j)$ where the swap rate is by definition constant. More interesting is the floating part.

Proposition 4.2.1. *The value of the LIBOR leg of a swap with constant notional amount is equal to the notional times the difference of the first and last discount factor. The PV of a floating rate note is equal to the notional.*

If the first-time discount factor is 1, i.e. $D(0,0) = 1$, the PV of the floating leg equals

$$\mathrm{PV(Float)} = (1 - p(0,T))N \ .$$

To prove this, we price a **Floating Rate Note (FRN)**, i.e. a fixed maturity bond with LIBOR-floating coupons. We set $L_j = L(t_{j-1}, t_j)$ for the LIBOR forward rate fixed at t_{j-1} with payment at t_j. The cash flows of the floating leg are not known at time 0 except the first one L_0. Although the cash flows are random, the claim is that the PV of the floating rate note is deterministic. To prove this, we replicate the random cash flow. Consider the period where we have to replicate the payoff $L_j = L(t_{j-1}, t_j)$, i.e. to obtain at t_j the payoff L_j. We need one unit of a currency at time t_{j-1}, which can be invested at the rate L_j such that we get in t_j the payoff $1 + L_j$. To obtain this we buy a zero $p(0, t_{j-1})$ and sell a zero $p(0, t_j)$. The long bond picks up the rate L_j at t_{j-1} and the balance of both bonds at t_j is L_j - i.e. the replication is accomplished. Consider the next cash flow L_{j+1}. Then the short bond $p(0, t_j)$ for the former cash flow will enter also as a long bond: He cancels. Therefore, considering a series of cash flows L_j, the replicating bonds cancel but the last one. This proves:

$$\mathrm{FRN}(0) = N$$

and hence $\mathrm{PV(Float)} = (1 - p(0,T))N$ for the floating leg of the swap.

Consider a 2y FRN with reset date each 6m, with notional $1'000$ and given spot rates. Setting the day count fraction to 1/2 for simplicity we get the following values, see 4.2:

Maturity	Spot Rate	Forward Rate	Cash Flow FRN	PV FRN
0.5	2.45%	2.45%	12.25	-
1	2.62%	2.76%	13.78	-
1.5	2.80%	3.08%	15.40	-
2	3%	3.45%	1017.37	-
-	-	-	-	1000

Table 4.2: Valuation of a FRN. The forward rates are calculated using simple compounding $F(0,S,T) = \frac{\frac{1+T\times R(0,T)}{1+S\times R(0,S)}-1}{T-S}$. The FRN cash flows are derived from $\text{CF}(T) = 1'000 \times \frac{F(0,S,T)}{2}$ and the PV follow from $\text{PV}(\text{CF}(T)) = \frac{\text{CF}(T)}{1+T\times R(0,T)}$.

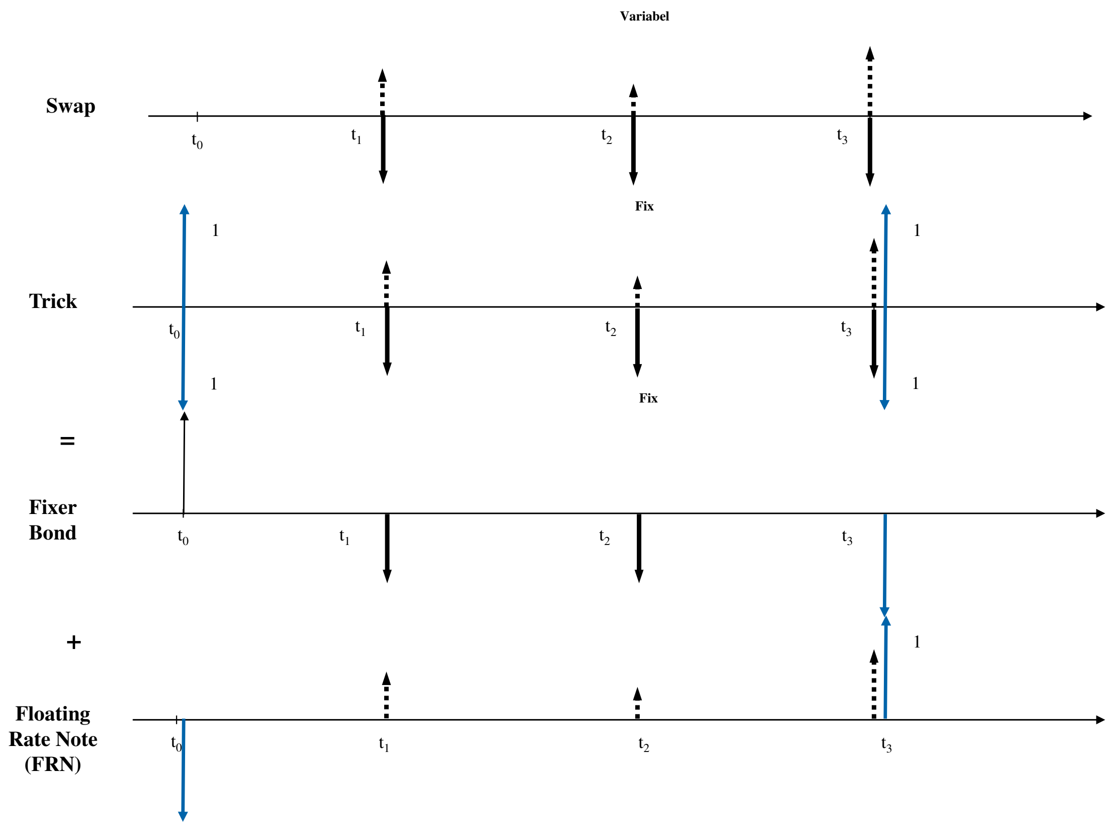


Figure 4.3: Graphical representation of a payer swap replication (payer means the party which pays the fixed rate and obtains the floating one). Dotted lines represents floating cash flows. Replication is obtained by virtually adding and subtracting notional amounts at the beginning and maturity of the swap. We assume for simplicity the same periodicity for the floating and fixed leg. The figure shows an important property of **risk structuring**: To obtain the cash flow profile of a new product one can add to an existing profile new products and add them vertically.

4.3 IRS and Forward Rate Agreements

IRS can be considered as a sequence of **forward rate agreements (FRAs)**: A FRA is a swap with a single floating and single fixed leg. We discuss the rationale for FRAs first and then the relation to swap pricing. Consider a client which would like to obtain a loan of CHF 10 Mio. starting in 6m with 6m maturity. The terms are fixed in 6m, i.e. LIBOR spot $L(6m, 6m)$ is the rate. The client believes that 6m LIBOR starting in 6m will be higher than present 6m LIBOR. He would like to freeze the loan term on the actual interest rate level, i.e. on the fixed forward LIBOR rate $K = F(0, 6m, 6m)$: The clients wants to swap the floating rate $L(6m, 6m)$ against the fixed rate $F(0, 6m, 6m)$. A FRA contract achieves this client's need:

- At initiation time 0 no cash flows are exchanged.
- At time 12m the client has to pay CHF $-10(1+L(6m, 6m))$ Mio. without an FRA. But the client would like to pay CHF $-10(1 + F(0, 6m, 6m))$ Mio.
- To achieve this he buys a FRA. Such a contract pays/receives an amount A in 6m and $A(1 + F(6m, 6m))$ in 12m such that A balances the payments in 12m between the unwanted risky payment without a FRA and the wanted fixed payment: A solves in **12m** the equation:

$$\underbrace{A(1 + L(6m, 6m))}_{\text{Balance}} - \underbrace{10(1 + L(6m, 6m))}_{\text{Without FRA}} = \underbrace{-10(1 + F(0, 6m, 6m))}_{\text{Desired Payment}} .$$

Solving for A, inserting the year-fraction $\alpha = \frac{\text{Hedging period}}{360}$ and $K = L(0, 6m, 6m)$ we get:

$$A = \frac{10\alpha(L(6m, 6m) - K)}{1 + \alpha L(6m, 6m)} .$$

Choosing spot rates $R(0, 6m) = 6.5\%$, $R(0, 12m) = 7\%$ and considering three interest rate scenarios A, B, C, we illustrate how the FRA provides an interest rate risk hedge. The no arbitrage relation between spot rates $R(s, t)$ and forward rates $F(s, t, u)$ (see notes after the example for details)

$$\left(1 + R(0, 6m)\frac{183}{360}\right)\left(1 + F(0, 6m, 6m)\frac{182}{360}\right) = \left(1 + R(0, 12m)\frac{365}{360}\right) \tag{4.2}$$

implies

$$K = F(0, 6m, 6m) = 7.26\%.$$

Note that a violation of (4.2) leads to a money machine. That is, assuming without loss of generality that the left hand side is larger than the right hand one, borrowing at the cheaper rate and investing at the higher one leads to a risk less profit.

The 6m LIBOR in 6m can

- be equal to the strike $F(0, 6m, 6m) = 7.26$ (scenario A),
- higher: $L(6m, 6m) = 8\%$ (scenario B);
- lower: $L(6m, 6m) = 7\%$ (scenario C) than the forward rate.

Table 4.3 summarizes the situation for the client.

Client Balance Sheet in Mio. CHF									
		0	0	6m	6m	6m	12m	12m	12m
	A	FRA	Loan	FRA	Loan	Total	FRA	Loan	Total
Scenario A	0	0	0	0	10	10	0	-10.37	-10.37
Scenario B	0.036	0	0	0.036	10	10.036	0.037	-10.404	-10.37
Scenario C	-0.013	0	0	-0.013	10	9.987	-0.013	-10.354	-10.37

Table 4.3: FRA under different scenarios. The figures are calculated as follows. For the payment A in Scenario B we have $0.036 = \frac{10(8\%-7.26\%)\frac{182}{360}}{1+8\%\frac{182}{360}}$. The loan at maturity is calculated as: $10.404 = 10 \times (1 + 0.08 \times 182/360)$.

This shows that with a FRA the client can completely hedge interest rate risk related to the loan. He pays in each scenario CHF -0.37 Mio. This corresponds to an interest payment on the loan using the forward rate of 7.26%, i.e. $10.37 = 10(1+0.0726*182/360)$.

In summary a FRA is based on three dates:

- Current time t,
- expiry time $S > t$,
- maturity time $T > S$.

FRA contracts allow to fix interest rates up to one year. Due to this short time to maturity they are classified as **money market** instruments. Instruments with a longer maturity are **called capital** market products.. They can be fixed 1 month up to two years in advance and the notional amount is not exchanged. The contract is settled two days before the starting date and the difference between the reference rate and the fixed forward rate is exchanged between the two parties. There are minimum sizes for FRA depending on the currency. In USD minimum sizes are one million or more.

Returning to swap pricing and using the no arbitrage relationship between zero bonds and forward rates we get

$$s_{0,T}(0) = \sum_{j=1}^{T} w_j L(0, T_{j-1}, T_j) \ , \ w_j = \frac{p(0, j)}{A_{0,T}(0)} \ .$$

The sum over all weights w_j equals 1. This shows that the IRS is a **weighted sum of FRA's**.

4.4 Constructing Time-Value of Money Curves

We construct the discount function starting from the par swap rates ('stripping the curve', 'bootstrapping'). Why is this meaningful? Discount factors are basic to value payoffs but they are not observable. We derive them using observable swap rates. This exercise is a **formal innovation**: Input are the par swap rates for maturities longer than 1 year.[3] We start with a $1y$ swap with the par swap rate $s_{0,1}(0)$ and $6m$ LIBOR for the floating leg. With Proposition 4.2.1 for the valuation of the floating leg and

$$\mathrm{Swap}_{\mathrm{Float}} = \mathrm{Swap}_{\mathrm{Fix}}$$

we get

$$N(D(0,1)-1) = s_{0,1}(0)D(0,1)\alpha_{0,1} \ .$$

Solving for the discounting factor as a function of the swap rate:

$$D(0,1) = \frac{1}{1+s_{0,1}(0)\alpha_{0,1}} \ .$$

To obtain $D(0,2)$ we have to consider a 2y swap with swap par rate $s_{0,2}(0)$. We have

$$N(D(0,1)-1) = s_{0,2}(0)\left(D(0,1)\alpha_{0,1} + D(0,2)\alpha_{1,2}\right) \ .$$

This shows that to determine $D(0,2)$, we need to know $D(0,1)$ (Bootstrapping, curve stripping). Solving, we have

$$D(0,2) = \frac{1 - s_{0,2}\alpha_{0,1}D(0,1)}{1+s_{0,2}\alpha_{1,2}} \ .$$

An immediate recursion gives

$$D(0,n) = \frac{1 - s_{0,n}(0)\sum\limits_{i=1}^{n-1}\alpha_{i-1,i}D(0,i)}{1+s_{0,n}(0)\ .\alpha_{n-1,n}} \ . \tag{4.3}$$

So far, we assumed that the discount factors or zero bonds as well as the forward rates are known. What if there are holes, i.e. times were no observable instrument exists? Then we have to construct the expressions, i.e. we have to **interpolate**. Such a construction should satisfy several requirements:

- **Mark-to-market.** The value of a dollar at a future date should be determined by liquid securities. This minimizes the risk that cash flows, are misspecified.
- **Stability.** The constructed term structures should be stable when switching from one structure to another one. Switching from a meaningful discount curve to a forward curve should also provide a meaningful forward curve.

[3] We do not consider shorter maturities here where we would use money market instruments.

Period	SARON	Period	LIBOR %	Period	Swiss Gov. Bonds
o/n	0.04				
		1m	0.13833		
		3m	0.17		
		6m	0.24		
		12m	0.52		
				2y	0.605
				3y	0.811
				4y	1.029
				5y	1.229
				7y	1.538
				8y	1.651
				10y	1.818
				20y	2.162
				30y	2.278

Table 4.4: CHF interest rates as of January 2011. SARON (Swiss Average Rate Overnight) is an overnight interest rates average referencing the Swiss Franc interbank repo market. *Source: Swiss National Bank.*

- **Smoothness.** Curves should not be ragged unless a sound economic explanation exists.
- **Consistency.** Estimated term structures today should be consistent with the dynamics of interest rate models. More precisely which parameterized families used to estimate the forward rate curve are consistent with arbitrage free interest rate models? We do not consider this issue and refer to Filipovic (2009).

We only discuss some basic ideas and do not consider advanced mathematical issues.

Different institutions generate different curves. The government term structure is generated using government bonds. The interbank term structure is constructed using money market instruments, futures and swaps, see Table 4.4. Finally the corporate term structure is constructed relative to a 'risk less' curves such as the government term structure. Consider the data for CHF in Table 4.4 using money and capital market instruments. This table shows that different instruments are used for different maturities and that for all cash flows not on the grid an **interpolation** procedure is necessary. The data in the table are **blended**: There is for example no overlap, i.e. for 3m maturity one could use LIBOR rates or prices of 3m futures contracts. Which one should one choose? The answer depends on several factors such as liquidity, economic considerations, market constraints. Suppose that a transaction is on-balance sheet and time value of money is constructed using off-balance sheet FRAs. This can turn out to be not a sound choice.

There are several methods to find a curve which interpolates the observed data.

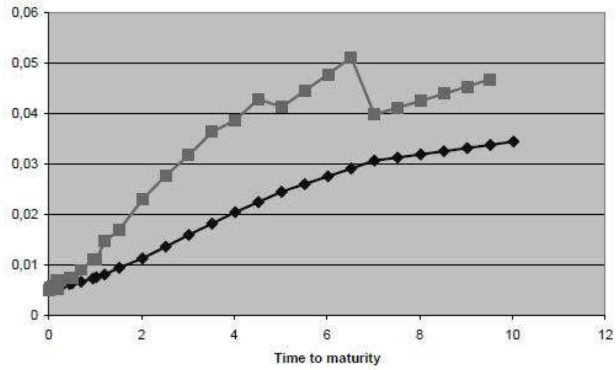


Figure 4.4: Smooth yield curve and the associated forward rate curve. The jagged forward curves arises in the swap region where linear interpolation is used. *Source: Filipovic (2009)*

The first one is to use a full cash flow view. This means that the vector of the n market instruments $\vec{X}$ (LIBOR, futures, swaps, etc.) is represented as

$$\vec{X} = C\vec{D} + \epsilon$$

with C the cash flow matrix and $\vec{D}$ the N discount factors for the searched term structure. The error term ϵ allows to treat statistically bid and ask differences or outliers in prices. One needs to transform all bond and money market instruments into this format, i.e. one builds a cash flow matrix C for the bonds, futures, LIBOR rate products and the swaps. The term structure $\vec{D}$ is found by minimizing the quadratic distance between $\vec{X}$ and $C\vec{D}$. This optimization has some serious drawbacks. First the number of market instruments n is much smaller than N since each cash flow of each instrument generates a column in the C matrix. This leads to an optimization problem with many possible solutions. Furthermore, the matrix C has many entries which are zero since most cash flows appear only once. There is nothing in this approach which prevents one to obtain a ragged term structure curve.

The next methods use parametric curve representations. The first guess is to linearly interpolate between existing rates. Suppose the zero rate curve is constructed in this way. Then the forward rate curve follows. Since the forward rate is basically a derivative of the zero rate curve[4] errors or kinks in the linear approximation lead to jagged

[4]The no arbitrage relation $1 + f(0,S,S+\Delta) = \frac{p(0,S+\Delta)}{p(0,S)}$, we get $f(0,S,S+\Delta) = \frac{p(0,S+\Delta)-p(0,S)}{p(0,S)} \sim$

forward rate curves, see Figure 4.4. The goal is to therefore to use higher order polynomials. Three approaches are the so-called B-splines or cubic splines, smoothing splines and exponential-polynomial (Nelson-Siegel, Svensson) approaches. We consider the last approach which is used by most central banks to construct the term structure. These approaches estimate the forward curve by minimizing the distance to bond prices. This is a non-linear optimization problem since the estimated forward curve is a sum of terms

$$\text{Polynomial} \times \text{Exponential function} .$$

The coefficients in the polynomial and in the exponential function are estimated by calibration in the non-linear optimization program to bond market prices. Figure 4.5 shows curves for the Nelson-Siegel family.

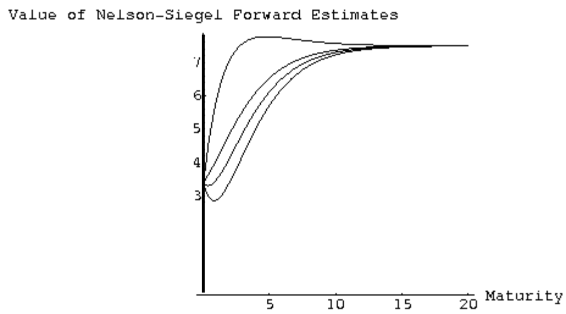


Figure 4.5: Smooth curves for the Nelson-Siegel family defined by $F = z_1 + (z_2 + z_3 T)e^{-z_4 T}$. The parameters z_1, z_2, z_4 are kept fix and different value for z_3 are used in the figure.

The figure shows that the curves F for the forward rate estimate is smooth and flattens out towards the long end of maturity. The approach is flexible in choosing the degree of the polynomial terms and it is consistent with the dynamics of the chosen interest rate model. For T to infinity the estimated curve becomes equal to the constant z_1 which is interpreted as the long term rate. The second term z_2 represents the short term rate due to the exponential decay and z_3 is the intermediate rate term representation.

We illustrate the bootstrapping method for the construction of a government term structure using the data given in Table 4.5. To obtain the interest rate for 14m using

$\frac{\partial p(0,S)}{\partial S}$.

Period	Zero Coupon Rate
o/n	3%
1m	3.1%
2m	3.2%
3m	3.3%
6m	3.5%
9m	3.9%
12m	4%
Maturity	Bond
14m	3.5% coupon, price 102
21m	5% coupon, price 100.5
22m	4.5% coupon, price 99

Table 4.5: Hypothetical zero coupon rates for a government term structure.

the 2m rate one solves:

$$102 = \frac{3.5}{(1+0.032)^{(1/6)}} + \frac{103.5}{(1+x)^{7/6}}$$

The result is the rate 4.318% for 14m. In the same way one obtains for 21m and 22m using the 9m and 12m rates the zero rates 5.479% and 5.062% which shows that the zero rate curve is inverse after 12m.

Zero coupon rates at intermediate dates can be obtained by linear interpolation of known rates. We noted that this approach often leads to jagged forward rate term structures. Interpolation by higher order polynomials is more adequate. We consider interpolation by 3rd order polynomials, i.e. the spot rate $R(0,t)$ is given by

$$R(0,T) = at^3 + bt^2 + ct + d$$

with the constraint that the curve matches known rates at dates t_1, t_2, and so on. Suppose that we are given 4 rates at the dates

- $t_1 = 1y, t_2 = 2y, t_3 = 3y, t_4 = 4y$ with
- values $4, 4.5, 5, 5.3$ percent, respectively.

Using the cubic interpolation we search for the intermediate rate $R(0, 2.5y)$, i.e.

$$R(0, 2.5) = a(2.5)^3 + b(2.5)^2 + c(2.5) + d = ? \ .$$

The condition that this unknown rate matches the 4 given ones is equivalent to a linear system:

$$Mx = y \ , \ x = (a,b,c,d)' \ , \ y = (4\%, 4.5\%, 5\%, 5.3\%,)' \ , \ M = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 8 & 4 & 2 & 1 \\ 27 & 9 & 3 & 1 \\ 64 & 16 & 4 & 1 \end{pmatrix}$$

where the matrix M has the time index powers as entries. Then $x = M^{-1}y$ provides the unknown parameters and gives $R(0, 2.5)$. The inverse matrix is:

$$M^{-1} = \begin{pmatrix} -1/6 & 1/2 & -1/2 & 1/6 \\ 3/2 & -4 & 7/2 & -1 \\ -13/3 & 19/2 & -7 & 11/6 \\ 4 & -6 & 4 & -1 \end{pmatrix} .$$

This implies for the vector

$$x = (-0.00033, 0.002, 0.00133, 0.037)$$

which gives

$$R(0, 2.5) = -0.00033(2.5)^3 + 0.002(2.5)^2 + 0.00133(2.5) + 0.037 = 4.762\% .$$

We next consider interpolation using **exponential weighting**. Consider the discount factors in Table 4.6. To calculate the par rate at $a = 1.5$ months we search for a discount

Grid	O/N	1w	1m	2m	3m	6m	1y	2y
Date	18/5/90	24/5/90	18/6/90	17/7/90	17/8/90	19/11/90	17/5/91	...
Par Rate	14.25	14.75	15	15.125	15.15625	15.375	15.55	...

Table 4.6: Par Rates of the British pound.

factor for a intermediate date a of the form

$$D(0, a) = e^{-R_a(t_a - t_0)} ,$$

where $t_1 < a < t_2$. The unknown rate R_a is the weighted average

$$R_a = \lambda R_1 + (1 - \lambda)R_2$$

of the known factors. The weight $\lambda \in (0, 1)$ is given by the relative position of the date t_a between t_1 and t_2, i.e.

$$\lambda = \frac{t_2 - t_a}{t_2 - t_1} .$$

We get

$$D(0, a) = D(0, t_1)^{\frac{\lambda(t_a - t_0)}{t_1 - t_0}} D(0, t_2)^{\frac{(1-\lambda)(t_a - t_0)}{t_2 - t_0}} .$$

Using this approach we search the discounting factor for 6, June 1990. The date lies between the 1m and 2m dates in the Table 4.6. The par rates are transformed into discounting factors as:

$$D(0, 6m) = \frac{1}{1 + \text{Par Rate}_{6m}\alpha_{0,6m}} = 0.9273 .$$

There are 31 days between the calendar date for the 2m and 1m factors and the weighting factor λ is equal to 4. This gives the intermediate discounting factor:

$$D(0, 20/6/90) = 0.9870^{\frac{34\times 27}{31\times 30}} 0.9753^{\frac{34\times 4}{31\times 61}} = 0.9854 \ .$$

We apply the methods to **swap pricing**. Table 4.7 shows how different rates are derived from the given swap rates. We note:

- The forward curve is a function $t \to F(t, T)$.
- The zero or discount curve is a function $T \to p(0, T)$
- The par swap rate curve is vector of spot starting swap rates for all maturities.

Maturity	Swap Rate	Discount Factor	Spot Rates	Forward Rates
1y	4.50%	0.95638	4.5615%	-
2y	4.95%	0.90647	5.0324%	5.55%
3y	5.39%	0.85158	5.5015%	6.57%
4y	5.57%	0.80151	5.6872%	6.28%
5y	5.68%	0.75409	5.8071%	6.31%

Table 4.7: Rates and their transformation. To obtain the discount factors from the swap rate we use (4.3). To get the spot rates from the discount factor we use $R(0,T) = \left(\frac{1}{D(0,t)}\right)^{1/T} - 1$ and the forward rates are calculated as $F(0,S,T) = \frac{D(0,T)^T}{D(0,S)^S}$. The day-count factor reads act/360/100 =1/36000*365=0.0101388.

Using these rates we price a non-forward starting $5y$ swap with a notional of 50 Mio. in a given currency. Table 4.8 summarizes the floating leg pricing. The PV of the floating leg is $-12'395'159$. This is one method of pricing. We could use the trick by inserting the notional amounts, see Proposition 4.2.1. We then immediately get

	1y	2y	3y	4y	5y
Rates	4.5615%	5.5518%	6.5750%	6.2827%	6.3128%
Cash flows	-2'280'743	-2'775'921	-3'287'486	-3'141'361	-3'156'409
PV of cash flows	-2'181'246	-2'516'291	-2'799'551	-2'517'843	-2'380'228

Table 4.8: Floating leg pricing. Up to 1y spot rates are used for longer maturities forward rates apply.

$$-12'395'159 = -50'000'000(1 - 0.75409) \ .$$

We price the fixed leg using 1% as an *ad hoc* fixed rate. The result is given in Table 4.9. The PV using 1% fixed is $2'135'015$. To make the value of the swap zero at time 0,

	1y	2y	3y	4y	5y
Fix 1%	1%	1%	1%	1%	1%
Cash flows	500'000	500'000	500'000	500'000	500'000
PV of cash flows	478'188	453'235	425'789	400'757	377'047

Table 4.9: Pricing with a fixed 1% rate.

the fixed swap rate s follows:

$$s = -\frac{\text{PV}_{\text{Floating}}(0)}{\text{PV}_{\text{fix at }1\%}(0)} = 5.806 \ .$$

We consider the **pricing of the IBM-Worldbank Swap**. IBM had to make the following payments:

- 12'375 million CHF p.a. on March 30 from 1982 through 1986 when it also had to repay the principal of CHF 200 Mio.
- 30 million DM p.a. on the same dates and the principal payment was DM 300 Mio.

Through the swap, IBM wanted to receive these payments from the World Bank. What are the equivalent USD payment for IBM to the above World Bank payments? We need the present value of the foreign currency payments the World Bank had promised to make. Suppose a flat term structure with the interest rates at transaction settlement date August 11, 1981 of 8 percent (CHF)and 11 percent (DM). The settlement date of the swap was August 25, 1981. We then have for the PV of the fixed CHF payments

$$\text{PV}_{\text{fix, CHF in CHF}} = 191'367'478 = \frac{12'375'000}{(1+0.08)^x} + \sum_{i=1}^{3} \frac{12'375'000}{(1+0.08)^{i+x}} + \frac{212'375'000}{1+0.08)^{4+x}}$$

as of August 25, 1981. Here x are 215 days instead of 360 days for the subsequent years since the first payments were due March 30, 1982. The terms of the swap were agreed to weeks before i.e. August 11. Using the 2w forward contracts for August 25 for conversion where 1 USD was worth 2.18 CHF, we get

$$\text{PV}_{\text{fix, CHF in USD}} = 87'783'247$$

as of August 25, 1981. With the same procedure one gets:

$$\text{PV}_{\text{fix, DM in USD}} = 117'703'153$$

as of August 25, 1981 (with FX of 2.56). The total USD payments of the World Bank were then $Z = 205'485'000$. The World Bank could borrow a present value of USD Z with a payment schedule matching the payment schedule of the IBM debt. The World Bank had to issue the debt and had to pay commissions and expenses of 2.15% of par. Consequently, to get a net present value of Z it could issue debt at par for USD $210'000'000$ with a coupon of 16 percent. The World Bank would receive 97.85% of USD $210'000'000$, i.e. which amounts to Z.

4.5 Swaps and Asset Liability Management (ALM)

The treasury of bank is responsible for interest rate risk of the bank's balance sheet. We consider a simple balance sheet in order to understand some aspects of the risk:

- 1 asset A with price 100, for example a loan.
- 1 liability L with price 60.
- The difference 40 is the capital E.

We assume a duration of $D_A = 5y$ for the asset, an duration of $D_L = 17.7y$ for the liability. Hence, an interest rate increase of 1 percent leads to a loss of 5 percent on the assets. Assume that the treasurer fears that the flat interest rate curve falls from actual 6 to new 5 percent. We assume the same rates for the assets and liabilities.

Was is the impact of such a change on the capital? Using the duration concept we get for the asset side (with P_A the price of the asset):

$$P_A(5\%) - P_A(6\%) = -\frac{1}{1+r} D_A \times \delta r \times P_A(6\%) = \frac{1}{1+0.06} 5 \times (-0.01) \times 500 = +23.6 \ .$$

Similarly, for the liability:

$$P_L(5\%) - P_L(6\%) = -\frac{1}{1+0.06} 17.7 \times (-0.01) \times 400 = +66.6 \ .$$

This induces a change on the capital of

$$100 + 23.6 - 66.8 = 56.8 \ ,$$

i.e. a loss follows if the treasurer applies the duration concept **independent** on the asset and liability side.
But the goal of the treasurer is that the duration of capital D_E remains unchanged if the interest rate change happens. To achieve this, one first needs to know D_E. We have from the balance sheet and the linearity of the duration concept:

$$P_E(r + \delta r) - P_E(r) = P_A(r + \delta r) - P_A(r) - (P_L(r + \delta r) - P_L(r))$$

i.e.

$$P_E(r + \delta r) - P_E(r) = -\frac{1}{1+r} D_A \delta r P_A(r) + \frac{1}{1+r} D_L \delta r P_L(r) \ .$$

From the definition of the duration we have on the other hand

$$P_E(r + \delta r) - P_E(r) = -\frac{1}{1+r} D_E \delta r P_E(r) \ .$$

Equalizing the expressions, we get

$$D_E = \frac{P_A}{P_A - P_L} D_A - \frac{P_L}{P_A - P_L} D_L \ ,$$

i.e. the duration of the capital is equal to the weighted sum of the asset and liability duration. We get with our figures

$$D_E = -45.8$$

i.e. the duration is strongly negative. To achieve neutrality $D_E = 0$ we need

$$P_A D_A = P_L D_L \ .$$

Hence, the treasurer needs a financial product

- with a strong positive duration,
- which has a PV of zero, i.e. does not change the present value of the balance sheet.

In order that a product has a PV of zero, it needs two CF streams, i.e. a swap. He enters into a contract where he sells the fixed part and gets the floating one since the floating one with 3m LIBOR has a duration of 3m and the fixed one has much longer duration. He exchanges long duration against short term duration. This makes IRS a favorite product for treasurers. This is a payer swap, i.e. the buyer of the swap pays the fixed CF.

4.6 Total Return Swaps

The concept of a swap is much broader than the above IRS examples. Basically any transaction which exchanges cash flows or assets is a swap. One can exchange currencies, commodities, stocks, and so on. The exchange may remain within one asset class but can also cross them, such as cross currency swaps. Third, swaps can be used to structure retail structured products.

We consider **Total Return Swap (TRS)**. A TRS swaps the total return of a single asset or basket of assets in exchange for periodic cash flows, typically a floating rate such as LIBOR plus/minus a basis point spread and a guarantee against any capital losses. Contrary to a plain vanilla swap in a TRS the total return, i.e. cash flows plus capital appreciation/depreciation, is exchanged, not just the cash flows. TRS do not transfer actual ownership of the assets, as in a repurchase agreement (Repo), i.e. the sale of securities together with an agreement for the seller to buy back the securities at a later date. A TRS allows for greater flexibility, reduces up-front capital and allows for higher leveraging. Market Participants include investment banks, commercial banks, funds (mutual, hedge, private equity, pension), insurance companies, governments, non-governmental (NGO) organizations and treasury departments of large multinational corporations.

The two parties in TRS are the Total Return Payer (TRP) and the Total Return Receiver (TRR). The TRP pays the total return from a reference asset and receives LIBOR plus a spread. The TRR receives the total return and pays LIBOR plus a spread. TRS are (still) off-balance sheet transactions, i.e. the TRR obtains the cash flows of a reference asset without actually owning it. Figure 4.6 shows the diagram of a TRS.

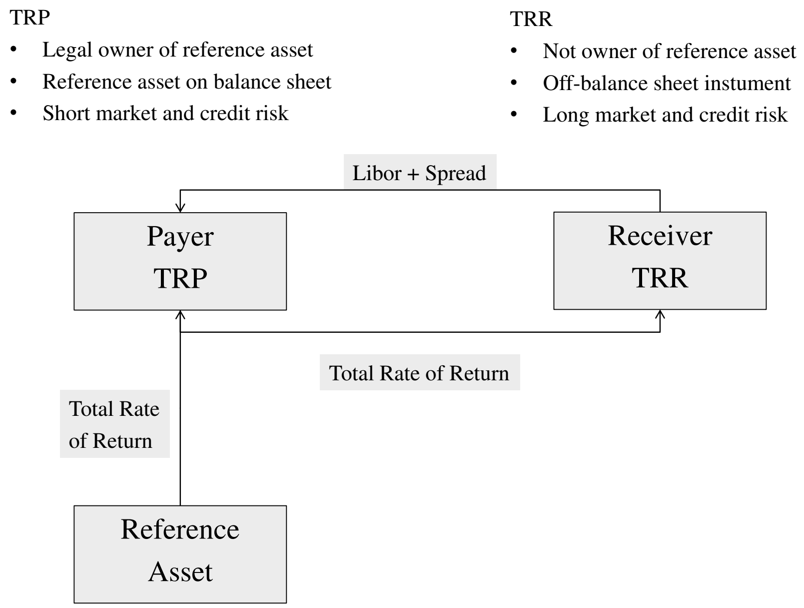


Figure 4.6: Total Return Swap (TRS).

If the reference asset defaults, the TRR bears the burden of the default. The TRR must then either pay the TRP the difference of the reference asset at inception of the swap and at the time of default or take delivery of the defaulted asset and pay the TRP the price of the reference asset at inception of the swap. The TRP has exchanged credit risk of the reference asset for credit risk of the TRR. The TRP demands collateral from the TRP to mitigate credit risk.

Hedge funds and special purpose vehicles (SPVs) used TRS for 'leveraged balance sheet arbitrage'. A hedge fund wants an exposure to a particular asset to earn high asset returns but do not wants to raise the capital to buy the assets. 'Leasing' them via a TRS is less costly. The bank generates additional cash flow by charging a spread above the market returns it receives from lending and receives a guarantee against depreciation of the assets. The synthetically transferred asset can range from single stocks to complex derivatives.

A TRS somewhat resembles a Credit Default Swap (CDS). Both a TRS and CDS are **off-balance** sheet and relate to the potential for default if the TRS reference asset is a loan or bond. But while a CDS only protects against loss caused by specific credit default events a TRS protects also against creditworthiness migrations. TRS has non-standard terms, whereas a single name CDS normally follows standard ISDA documentation. The greatest benefit and risk of a TRS is leverage. This means that up-front capital to execute a trade can kept low. This makes TRS a favorite of hedge funds. Other possible benefits are lower financing costs using a TRS. Such benefits exists if the same risks are priced different in different markets. Figure 4.7 shows the impact of leverage from TRS. The difference between a direct or cash investor in the reference asset and a hedge funds investing in the TRS is shown. If the funding costs of the hedge funds largely increase, in the example above 7 percent, then there will be a negative net swap spread and the return generating leverage become a bad for the hedge fund. Typically due to the high leverage of hedge funds such a situation leads to bankruptcy of the vehicle.

	Comments	HF A	HF B	Cash Investor
Libor		4%	4%	
Ref. Asset Yield	Libor + 300 bps	7%	7%	7%
Funding Cost	Libror + 100 bps for HF	5%	5%	
Net Swap Spread		2%	2%	
Collateral Amount		5%	10%	
Leverage		20 to 1	10 to 1	1 to 1
Leverage Swap Return		40%	20%	
Net Interest on Funding		1%	1%	5%
Net Return		39%	19%	2%

Figure 4.7: Leverage in TRS investment for two hedge funds. The figure compare these two investors with a cash investor. If funding costs exceed the reference asset yield, say they are up to 8 percent, then also the losses react to leverage. Using a leverage factor of 20 the losses are 20 percent in this scenario.

The investment return risk is born by the TRP. While the Total Return Payer retains the reference asset on its balance sheet, the Total Return Receiver assumes the risk of capital losses by making guarantee payments to the TRP that offset any drop in asset value. Both parties face interest rate risk in a TRS. Interest Rate Risk is typically higher to the investor who does not necessarily have direct access to LIBOR financing, whereas the TRP does. Liquidity risk may exist if the TRS terms specify physical delivery of assets between the parties. For example, if the TRS requires the TRP to deliver specific high yield bonds at expiration, and these bonds defaulted during the life of the TRS, it may be difficult to acquire them at reasonable valuation in the open market if the bank does not have them in its inventory. Counter party risk is a significant factor in certain transactions. If a hedge fund makes multiple TRS investments in similar assets, any significant drop in the value of those assets would leave the fund in a position of making ongoing coupon payments plus capital loss payments against reduced or terminated returns from the asset. Since most swaps are executed on large notional amounts between 10 and 100 Mio. USD, this could put the TRP at risk of a hedge fund's default if the fund is not sufficiently capitalized. Hedge fund counter party risk is accentuated due to secrecy and minimal or nonexistent balance sheet reporting obligations. Bankruptcy risk may exist where the reference asset is a single large capital asset, such as a industrial building mortgage or airplane loan. If the borrower defaults or files for bankruptcy, the loan payments may terminate, effectively eliminating the asset returns to the Total Return Receiver and requiring large capital loss payments to the TRP.

4.7 Documentation - ISDA Agreements

We introduce to the ISDA agreements, i.e. the **innovation of documentation standards**. These agreements apply to Over-the-Counter (OTC) derivatives, i.e. derivatives which are not exchange traded and which are tailored to the specific requirements of customers, see Figure 4.8 for an overview of **derivatives**.

Figure 4.9 shows the size of OTC derivative markets world wide for different risk factors.

- The OTC markets are much larger in size than their stock exchange counter parts. OTC markets face a strong change in regulation and transparency requirements, see the Section Financial Crisis.

- The Notional Amount and Gross Positive and Negative Market Values are shown in Figure 4.9. The first one is significantly larger than the second ones. Notional amounts are pure size measures. They are weak indicators for risk figures since risk of derivatives depends on volatility, correlation, interest rates and other parameters. The gross market value is equal to the sum of absolute value all contracts with a positive or negative replacement value. 'Gross' means that contracts with the same counter party are not netted, nor is a positive and negative replacement value within the same risk category netted. Replacement values reflect the marked to market values of all receivables of the open derivative positions of clients and proprietary positions of the banks at fixed date. That is, a positive replacement value is the non-defaulting party would receive if the contract were to be terminated today. Positive replacement values are assets of a bank and negative ones are liabilities.

- The column Gross Credit Exposure summarizes all gross values of contracts with a positive replacement value after all allowed netting operations. Netting implies a:

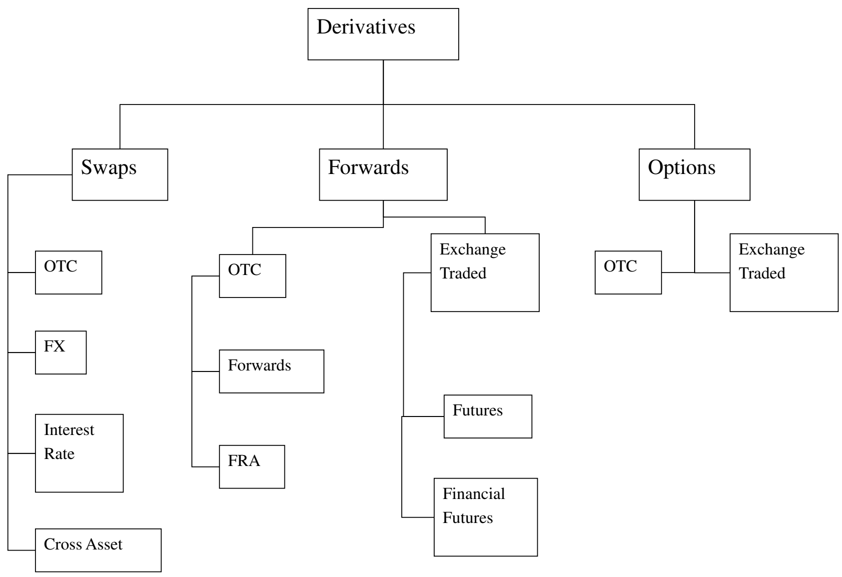


Figure 4.8: Derivatives classification. Warrants and Structured Products are not contained in this list. Swaps and forwards are called linear derivative contracts, options are non-linear ones.

  - Reduction of credit risk
  - Reduction of settlement risk
  - Reduction of liquidity risk
  - Reduction of systemic risk

- The figures are classified according to the risk factors: FX, Interest Rates, Equity, Commodities, Credit Risk in the form of Credit Default Swaps (CDS). Within each class non-conditional derivatives (Forwards, Swaps) and conditional derivatives (Options), credit derivatives on single-name and on an index of several names are distinguished.

The documentation of OTC derivatives uses standard forms of documentation developed by the International Swaps and Derivatives Association (ISDA) which represents more than 700 Participants in the OTC derivatives market. The documentation consists of Master Agreement, Definitions, Templates and User's Guides. The goal of the Master Agreement is to document bilateral derivative transactions in a standard form. The Master Agreement consists of several parts: a body, a schedule, transaction confirmation and optionally a credit support annex (CSA). A major concern of ISDA is counter party risk, i.e. how can this risk mitigated and how can this risk properly priced in bilateral derivative transactions? Credit risk of derivative transactions such as FRA and swaps differs from credit risk of bonds or loans. For linear derivatives in general no funds are exchanged between the parties at initiation. Therefore, the notional amount is a poor measure of credit risk, i.e. the market value of the derivatives is in general much smaller than the notional amount of its underlying. On the other side the derivative contract depends on an underlying value and large fluctuations during the life time, i.e. the future exposure is different from current exposure.

If a counter party defaults prior to settlement of the contract the non-defaulting party will try to replace the defaulted contract with an economic new contract. If the market value is positive, the loss is equal to the replacement costs. Counter parties possess many bilateral contracts among each other. In case of default each contract is treated as a unique legal contract unless there is a **netting provision** specified in the Master Agreement. Netting means that the gross claim between counter parties is replaced by a single net claim. Two types of netting are of particular importance: **Payoff netting** and **close-out netting**, see Figure 4.10.

Payment netting means that positions in the same currency and the same date are offset. This mechanism reduces settlement risk of derivative transactions. Close-out netting refers to a bilateral arrangement where both parties agree to terminate all obligations, i.e. even if they are not yet due, if default or another termination event occurs. The gross market value is added up and a single payment is obtained by the party with a negative net portfolio value. If there is at least one transaction with a positive market value then close-out netting reduces credit risk. To see this effect, let $E_t(i)$ be the exposure at time t of transaction i. Without any netting agreement the overall credit exposure between

Table 19: Amounts outstanding of over-the-counter (OTC) derivatives

By risk category and instrument

In billions of US dollars

	Notional amounts outstanding					Gross market values				
Risk Category / Instrument	**Dec 2008**	**Jun 2009**	**Dec 2009**	**Jun 2010**	**Dec 2010**	**Dec 2008**	**Jun 2009**	**Dec 2009**	**Jun 2010**	**Dec 2010**
Total contracts	**598,147**	**594,553**	**603,900**	**582,655**	**601,048**	**35,281**	**25,298**	**21,542**	**24,673**	**21,148**
Foreign exchange contracts	**50,042**	**48,732**	**49,181**	**53,125**	**57,798**	**4,084**	**2,470**	**2,070**	**2,524**	**2,482**
Forwards and forex swaps	24,494	23,105	23,129	25,625	28,434	1,830	870	683	925	886
Currency swaps	14,941	15,072	16,509	16,347	19,271	1,633	1,211	1,043	1,187	1,235
Options	10,608	10,555	9,543	11,153	10,092	621	389	344	411	362
Interest rate contracts	**432,657**	**437,228**	**449,875**	**451,831**	**465,260**	**20,087**	**15,478**	**14,020**	**17,533**	**14,608**
Forward rate agreements	41,561	46,812	51,779	56,242	51,587	165	130	80	81	206
Interest rate swaps	341,128	341,903	349,288	347,508	364,378	18,158	13,934	12,576	15,951	13,001
Options	49,968	48,513	48,808	48,081	49,295	1,764	1,414	1,364	1,501	1,401
Equity-linked contracts	**6,471**	**6,584**	**5,937**	**6,260**	**5,635**	**1,112**	**879**	**708**	**706**	**648**
Forwards and swaps	1,627	1,678	1,652	1,754	1,828	335	225	176	189	167
Options	4,844	4,906	4,285	4,506	3,807	777	654	532	518	480
Commodity contracts	**4,427**	**3,619**	**2,944**	**2,852**	**2,922**	**955**	**682**	**545**	**457**	**526**
Gold	395	425	423	417	396	65	43	48	44	47
Other commodities	4,032	3,194	2,521	2,434	2,525	890	638	497	413	479
Forwards and swaps	2,471	1,715	1,675	1,551	1,781					
Options	1,561	1,479	846	883	744					
Credit default swaps	**41,883**	**36,098**	**32,693**	**30,261**	**29,898**	**5,116**	**2,973**	**1,801**	**1,666**	**1,351**
Single-name instruments	25,740	24,165	21,917	18,494	18,145	3,263	1,950	1,243	993	884
Multi-name instruments	16,143	11,933	10,776	11,767	11,753	1,854	1,023	559	673	467
of which index products	...	...	...	7,500	7,476					
Unallocated	**62,667**	**62,291**	**63,270**	**38,327**	**39,536**	**3,927**	**2,816**	**2,398**	**1,788**	**1,532**
Memorandum Item:										
Gross Credit Exposure						5,005	3,744	3,521	3,578	3,342

Figure 4.9: OTC Derivatives Statistics. *Source: BIS, Basel, 2011.*

Since in option-type contracts the exposure is always positive for the option buyer and negative for the writer it is only the buyer who faces the risk of counter party default. If forward-type products such as swaps are considered, the exposure can be positive or negative, i.e. both parties face counter party risk.

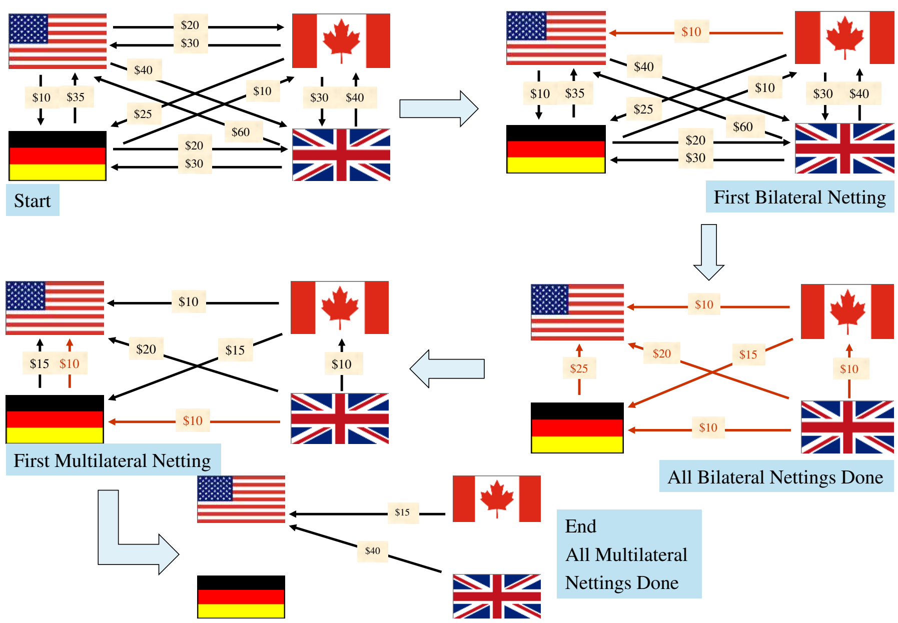


Figure 4.10: Netting examples. In the first transition a bilateral netting of USD 20 and USD 30 between a Canadian and an US counter party with an ISDA Agreement is performed. In the next chart all bilateral nettings are executed. In the next step multi lateral netting applies. This means that the liability of 25 USD form the German firm to the US one is split into a 15 and 10 USD flow. Since the German firm obtains 10 USD from the British firm, the British firm can directly hand the 10 USD to the US firm leaving a claim of 15 USD between the German and US firm and no claim between the German and British firm. This procedure is continued leading to the final two claims.*Source:McGraw-Hill Companies, Inc., 2001.*

two parties is

$$\sum_i \max(0, E_t(i)) \ .$$

With close-out netting we have $\max(0, \sum_i E_t(i))$ and

$$\max(0, \sum_i E_t(i)) \leq \sum_i \max(0, E_t(i))$$

shows the credit risk exposure reduction due to close-out netting. This is present time view.

What is the exposure at a **future date**?

The positive value of a contract with value V_t at time t is the **e**xposure E_t:

$$E_t = \max(V_t, 0).$$

This is the loss in case of default for a bank. Negative values possess no loss potential from the bank's view. The **P**otential **E**xposure $PE_{t|s}$ equals the 95%-quantile of the discounted exposure at time $t \geq s$:

$$\mathbb{P}\Big(D^{-1}(s,t)\, E_t \leq PE_{s|t}\Big) = 95\%.$$

with D the discount factor. The probability P depends on the risk factors. We assume that the value driving market parameters affect the future exposure up to time t in the same way they did up to time s, i.e. there is no regime switch. Maximal Potential Exposure (MPE) at s is then the maximum over all $PE_{s|t}$ with $s \geq t$:

$$MPE_s = \max_{t \geq s} PE_{t\cdot} = \max_{t \geq s} \Big\{ PE_{t|s} \Big| \mathbb{P}\Big(D^{-1}(s,t)\, E_s \leq PE_{t|s}\Big) = 95\% \Big\}.$$

The goal is to find a standardized calculation of MPE for **all** types of derivatives. That for one decomposes MPE in a sum of a Replacement Value (RV) and a Risk Premium (RP), see Figure **??**

$$\text{MPE} = \text{RV} \ + \ \text{RP}. \tag{4.4}$$

The risk premium is set equal to the maximum over all 95%-quantiles of the future exposures. We refer to the literature for technical details.

4.8 Capital and Money Markets

Swaps are one instrument of the **capital markets**. These markets are a part of **financial markets** which we compare with **real (goods) markets.** In the latter one, goods and services are traded. These objects are generated by the input factors labor and income. The production function acts on them and produces the output. But on the input factor capital another market acts which maps capital into capital - the financial markets. A

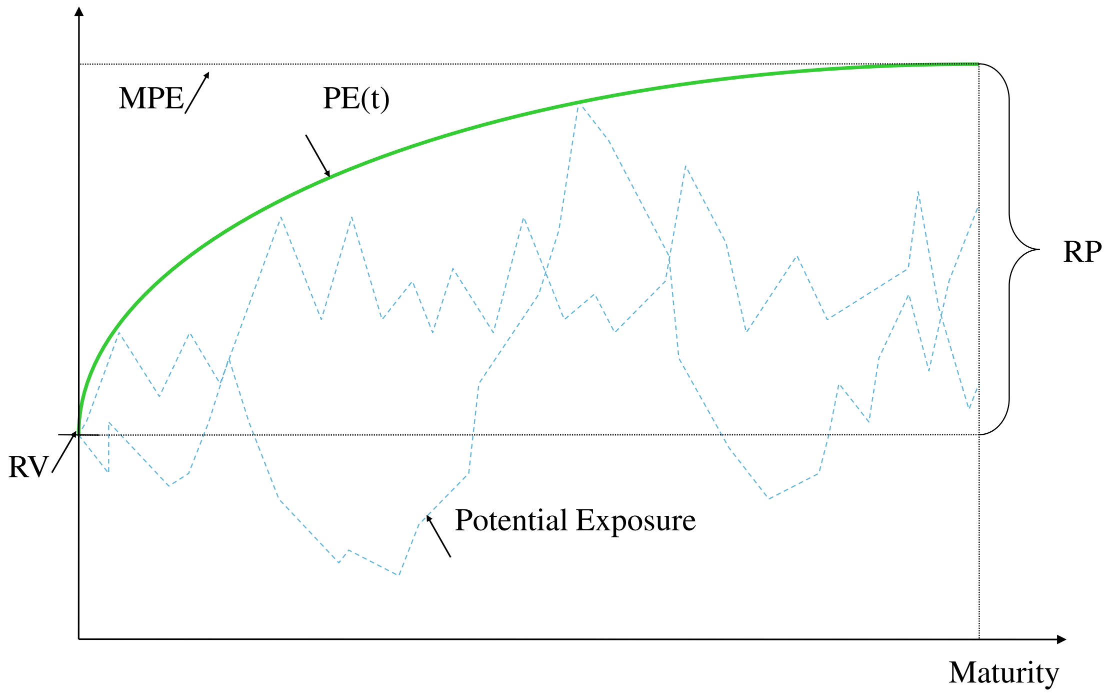


Figure 4.11: Illustration of the MPE decomposition.

second difference to the real markets is that financial markets are not presently oriented but they exchange future cash flows. A hybrid market are commodity markets were both spot and future markets exist. The financial market is in some sense acts on capital before the real market does: The financial markets allows to transform capital in the real market in the desired form: The input factor capital is transformed using financial markets. Derivative markets act on financial markets - they consider inputs of the financial markets and transform them into new financial markets products.Since there are different types of participants, different maturities and different legal setups the financial market is divided in sub markets:

- Money market,
- Capital and loan market,
- Foreign exchange market (FX),
- Derivatives market.

This division is not free of redundances. Forward rate agreements are derivatives and money market instruments. Roughly, money market instruments have a maturity up to one year and capital market ones of more than one year. On the loan market are traded all types of loans which are not traded in the capital market. Derivatives are financial instruments where the price depends on an underlying instrument, goods or reference rates (interest rates, indices). They are divided in:

- Contingent derivatives or option. In these contracts at least one party has the right (optionality) to do something, i.e. to buy say a stock at a prespecified price at a future date.
- Unconditional derivatives do not have any optionalities. Examples are FRA and swaps.

Contrary the so-called cash products - stocks, bonds, FX spot rates - derivatives have a time dimension, i.e. the contracts are about future prices. Another way to categorize derivatives is: Are they regularly traded at an exchange or are they OTC? Figure 4.12 shows a possible characterization of financial markets.

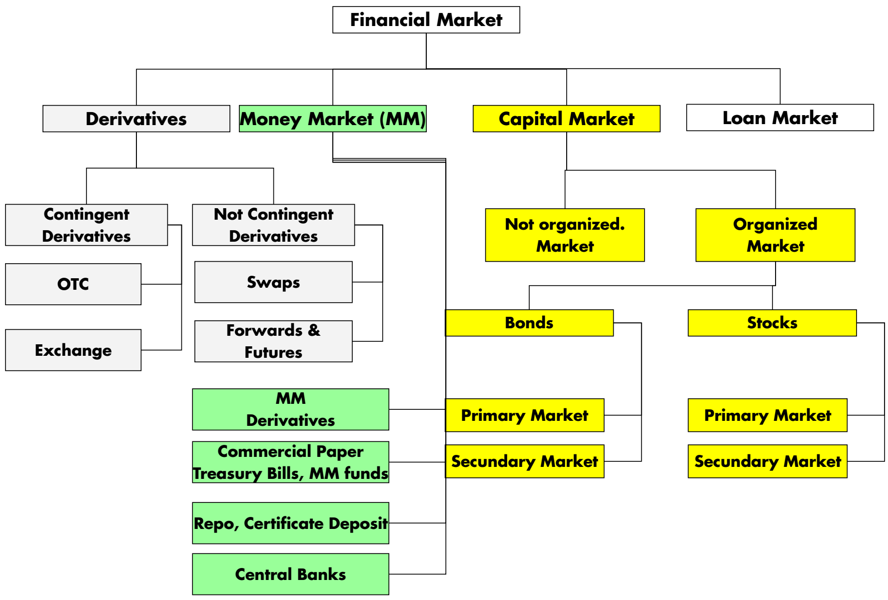


Figure 4.12: Financial markets. *Source: Wikipedia and own representation.*

The figure shows some products of the money markets (MM). These markets exist in almost all economies in the world. They are a first step towards the capital markets where maturities longer than 1 are traded. Actors in the MM are central banks, banks, governments, money market fonds and large corporates. The daily volume in CHF is in the two digit billion area. Figure 4.13 shows some figure of the MM.

The volume in the MM doubled in the 90s. The growth after 2000 until the start of the financial crisis 2007 was by a factor of seven. Before and after the outbreak of

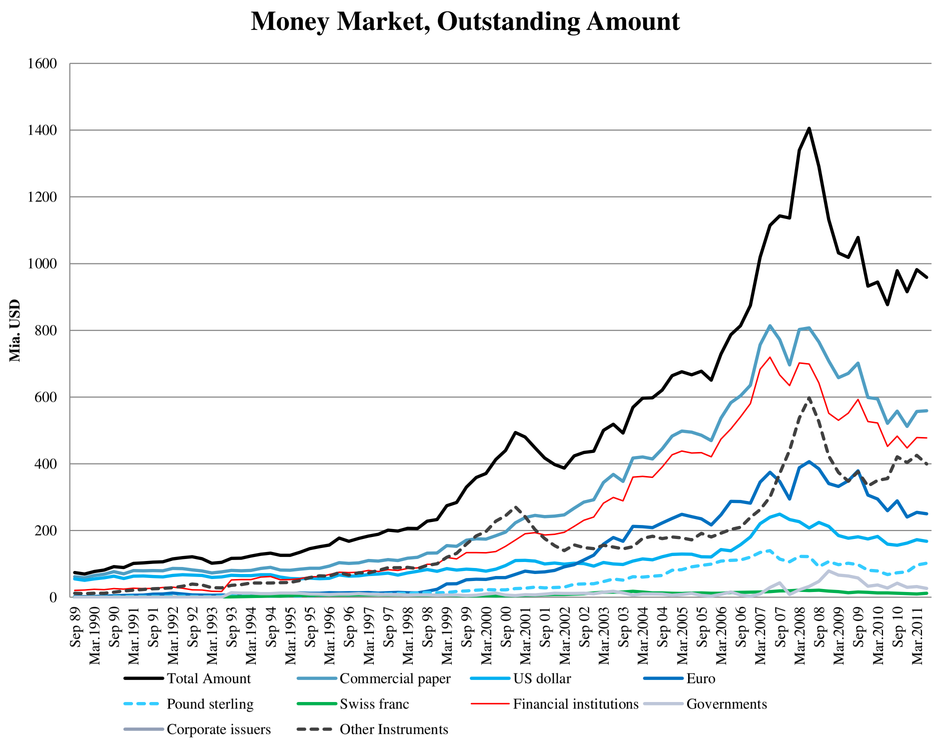


Figure 4.13: Money Market. *Source: BIS, 2011.*

the financial crisis both the growth and decline where more pronounced than before. The driver of these observation were financial institutions. The more marginal role of governments in this markets was only interrupted in autumn 2008. The money markets primary serves the banks to exchange liquidity, i.e. either to borrow or to invest excess capital. Large corporate use the MM to raise short term financing.

4.8.1 Secured Banking - Repo and SLB

An important instrument in MM are **Repo Transactions**. Reps business is part of **Prime Finance**:

- Lending and borrowing of securities, the (SLB) business.
- Repos, i.e. sale and repurchase agreements, are short term funding instruments.
- Synthetic Finance such as synthetic SLB or synthetic Repo. These transactions combine a SLB with a derivative where the underlying is the security of the SLB transaction.

Prime finance business changed a lot in the last years and is still in a transformation modus. There are several reasons:

- Transition to **collateralized banking**. Repo business can be considered as secured banking were collateral serves as a creditor protector for savings for non-retail investors. I.e. it is the counterpart to creditor protection on banking deposits. Creditors are bank, insurance companies, governments, firms or pension funds. Collateralization was always important in OTC transactions. The global financial crisis starting in 2008 showed that one has to fully understand credit risk and collateralization in OTC derivative transactions. Markets which are widely collateralized are for example fixed income repo, equity finance, exchange traded securities, OTC derivatives, securities lending, banks loans, asset backed securities. Collateral banking requires some specific functions within the bank. First, one needs a collateral management team. They run collateral operations, issue and receive margin calls, speak with clients or counter parties, provide services. The credit risk of the counter parties needs to be assessed and revalued by a credit analysis team. Sales are responsible for the onboarding process for new accounts and traders execute the trades. Collateral, in particular if it less liquid or complicated (exotic) needs to be valued by an independent valuation team. Other functions comprise legal, middle office, accounting functions. An important property of collateral is its eligibility, i.e. the extend how collateral can be converted into an economic value if the counter party defaults. Liquidity, quality in terms of embedded credit risk and the possibility to settle the collateral define the collateral eligibility. Cash is by far the most used collateral followed by government bonds, large-cap shares. Of lower eligibility are ETFs, mutual funds or guarantees.

- Cost reduction in the custody of securities. The administration of securities is costly. Lending and borrowing securities generates earnings which lower these costs.
- To cover short positions one has to borrow securities. Short positions can be the results of market making, the hedging of derivative positions or part of an investment strategy.
- Risk management. Collateral is an effective instrument to reduce credit risk.
- Increasing regulatory requirements. Risk weighted assets in the regulatory capital charge are reduced if one switches from unsecured to secured transactions.

A repo transaction can be cash or security driven. It is security driven if the investor wishes to lend a security. Repo transactions are bilateral trades between the Seller (security seller, i.e. loan receiver) and the Buyer, i.e. the security buyer or loan provider. These transactions can be described with two dates:

- At 0: Assignment of the securities from the Seller to the Buyer.
- At 1: Redemption of the loan and interest rate payments to the Buyer and reassignment of the security from the Buyer to the Seller.

The purchase price in 0 equal the market value (dirty price) of the underlying security minus an add on (Haircut). The haircut provides a restricted protection against falling security prices. The payback price equals the purchase price plus an agreed interest payment (repo rate), which depends upon the quality of the security. If the security losses value, a margin call follows. An economic incentive for a repo transaction for the Buyer is to obtain favorable rates compared to an unsecured loan and the Seller receives collateral.

Example:
While investors trade bonds on a stand alone basis, trading desks use repo jointly with bond trading: Buying a bond is completed immediately by selling the bond in a repo, i.e. one finances the bond. If a bond is sold short, the additional trade is a reverse repo. We consider an US Treasury Bond at the following dates:

- T trading day to buy the bond.
- $T_1 = T + 1$ settlement day for the bond. Start/opening the repo.
- $T_2 = T + 2$. Closing of the 1-day repo.

At the trading day T the trader buys the bond for the price $B(T)$ from a counter party A. The bond is settled at $T + 1$. **At** $T + 1$ the repo transaction starts to finance the bond. To achieve this

- the repo desk delivers the bond for 1 day, i.e. the period of the repo transaction is overnight from T_1 to T_2 for a price $B(T_1^{\text{Repo}})$, to the repo counter party and

- the repo desk agrees to buy the bond back at T_2 for the price

$$B(T_1^{\text{Repo}})(1 + r/360)$$

with r the **repo rate**.

The prices $B(T)$ and $B(T_1^{\text{Repo}})$ can differ at T_1. The difference is a residual cash position with a cash rate r_{cash}. This rate is in general different from the repo rate. **At** T_2 the repo desk pays the amount $B(T_1^{\text{Repo}})(1 + r/360)$ to the counter party, receives the bond back and delivers the bond for the price $B(T_1)$ to the buyer. The P& L of this transactions over 1 day has the following components:

$$\begin{aligned} \text{P\&L} &= P(T_1) - B(T) && \text{Price Change Bond} \\ &- B(T_1^{\text{Repo}})r/360 && \text{Repo Costs} \\ &+ (B(T_1^{\text{Repo}}) - B(T))r_{\text{Cash}}/360 && \text{Difference Repo vs. Cash Market.} \end{aligned} \tag{4.5}$$

We consider the following numbers:

- Notional 100 Mio. USD, coupon 4 percent.
- T is Oct 2 for trading the bond and settlement is Oct 3.
- Clean Price of the bond is 100'078'125 USD (= 100−02+ in US Treasury notation). with accrued interest the settlement price 100'110'911 follows. Accrued interest rate $\frac{3}{183} \times 0.04/2$, i.e. the bond accrues interest since Sept 30 and a half a year has 183 days.
- The repo rate r equals 3.4 percent, the cash rate is 3.5 percent.

Before the bond settles Oct 3, the repo desk finances the bond. The bond price changes from Oct 2 to Oct 3 by (100-05), i.e. the value of the position in dirty prices increased to

$$100'189'036 = (1 + 5/32 + \frac{3}{183} \times 0.04/2) \times 100\text{Mio. USD} .$$

Oct 3 the following payments/transactions are made:

- Bonds are received with value USD $100'110'911$ and exchanged for a secured loan of USD $100'189'036$ with the repo counter party.
- They deliver cash payments of $78'125$ USD.

Oct 4 the following payments/transactions are made:

- The repo counter party hands back the lent bond and obtains the repo rate interest:

$$100'198'499 = 100'189'036 \times (1 + 0.034/360) .$$

- The bond is sold from the repo desk to the buyer. The price equals the clean price of Oct 3 with Oct 4 settlement plus accrued interest. If the bond increased to 100-08, we have

$$100'293'715 = (1 + 8/32 + \frac{\mathbf{4}}{183} \times 0.04/2) \times 100 \text{ Mio. USD} .$$

The P&L components are:

- Change in bond price: $100'293'715 - 100'110'911 = +182'803$ USD.
- Repo costs : $-100'189'036 \times \frac{0.034}{360} = -9462$ USD.
- Difference Repo vs. Cash Market: $78'125 \times \frac{0.035}{360} = +7.7$ USD.

A 1-day P&L of 173'349 USD follows.

Contrary to the SLB business, repo is always of the type 'cash against security, whereas in SLB 'securities vs securities or cash' are exchanged. Both types face the same market risk. Settlement risk can be different and the rational can be either liquidity or security possession.

In fixed income repo one distinguishes between 'Repo Specials' and 'Repo General Collateral' (GC) markets. In Repo Special, similar to SLB, a specific security is considered which the buyer wants to possess to cover for example a short position in the security. In a GC Repo the rationale is borrow or invest money for a short time. The interest rate is derived from the actual MM rates and the quality of the collateral. If a repo is due one includes the interest into the payback price of the collateral. GC business is much larger than the Special one. Equity similar to FI repo. A main motivation is to finance equity or stock positions in the trading books or a short-term money market investment.

In a TriParty Repo a third party to the Buyer and Seller is responsible for administration and operations. The largest providers TriParty Repo programs are Clearstream and JP Morgan Chase which is large custodian bank (see Asset Management). TriParty Repos are designed for professional market participants. They allow to settle flexible baskets of securities as collateral in an efficient form. Repo transaction are used by central banks for the management of liquidity. Banks use repos to reduce credit risk, i.e. unsecured money market transactions are replaced by repos. For security dealers repo offer funding at favorable rates.

Eurex is a stock company hold in equal parts by SIX Swiss Exchange and Deutsche Börse AG. Eurex is one of the world wide largest exchange for futures and option trading. But Eurex also offers platforms for bond trading and for repo (Eurex Repo). This platform is open to all financial institutions. The Eurex Repo platform is a TriParty platform with integrated trading and settlement functionalities. Legal documentation is different to traditional OTC-Repo TriParty programs, since the Eurex platform integrates trading, settlement and legal documentation. Participants at Eurex Repo can choose from a broad menu of repo transactions. First, Eurex Repo operates in Swiss Franc and the Euro Repo Market. In the first market, the participants can trade directly in the Swiss interbank market as well as in the daily SNB repo auctions. An advantage of the Eurex Repo platform is that the securities which are received as collateral can be used immediately for a new repo transaction. This allows the banks to raise cash if they need to do so. Such a possibility does not exist in OTC-Repo transactions. The fraction of GC repo is 80 percent at Eurex-Repo. One expects that this amount will still further increase due to the increasing market standards for baskets which are not there for single securities: There are for example safety standards and term structures in GC business. The SNB-eligible baskets are a concrete example. The Eurex market consists of four links for the participants in CHF repos:

- Trading is via the Eurex Repo platform.
- Clearing, Settlement and Collateral Management takes place at SIX SIS.
- Cash Clearing is done via SIX Interbank Clearing.
- There is a link to SNB which publishes the SNB-eligible securities.

As an example, consider a bond trader (Seller) which wishes to borrow CHF 20 Mio. to finance for one week an investment of CHF 18 Mio. in Eidgenossen (Swiss Government Bonds) with 3 percent coupon. A repo buyer offers a repo rate of 2 percent. The seller accepts the rate. He delivers CHF 18 Mio. nominal against CHF 20 Mio. cash. At the same day he pays the buyer CHF 20 Mio. in exchange of the CHF 18 Mio. bonds. After one week the buyer gives back the bond to the seller. The seller pays back the loan of CHF 25 plus accrued interest:

$$20'000'000 \times \frac{0.02 \times 7}{360} = 7'777.8 \text{ CHF} .$$

The repo market played an important role in the financial crisis 2008. To understand this we discuss the difference between traditional and collateralized banking. The following discussion is based on Gorton und Metrick (2011). Traditional banking is characterized by a short chain of intermediation: Banks are in simplified way in between mortgage borrowers and depositors. Furthermore, depositors face a depositor protection. Such a protection serves as a buffer against bank runs. But this protection is limited in size for depositors and it also does not apply to non-retail clients. For them a possibility of protection for an investor is to obtain collateral from the bank using a repo. Hence, the repo rate corresponds to the interest rate in secured banking. The bank sells the collateral for a lower value - haircut - than its market value. If the collateral market value is CHF 1, the bank sells it for CHF 0.9 and buys it back for 0.95. This gives a repo rate of $5/95 \times 100 = 5.2$ percent and a haircut of 10 percent. Hence haircuts play the role of reserves in traditional banking.

The dependence between secured banking and the financial crisis follows from the value chain in loan banking. In traditional banking, it is the bank itself which decides about to whom a loan is given. In secured banking this function is outsourced. In the US creditors handed the loans or mortgages to the end users and the creditors immediately sold the mortgages to banks. The banks pooled the mortgages and defined new products for investor. This defines structured finance or securitization. Whereas in traditional banking the bank itself keeps the counter party risk of the borrower in secured banking this credit risk is passed from the direct creditor to the bank and from there via the newly structured product to the investors - which in the last crisis to a large extend also were banks. During the financial crisis there was no bank run on deposits in the US but a bank run on repo happened. To show this the authors consider the behavior of haircuts for securities in repo business: The construct a repo index. This index went up from low one digit percentage values before the crisis to almost 50 percent during the crisis. Some securities which would no longer to be considered as a collateral in effect faced a haircut of 100 percent. Such an increase of the collateral buffer in a hugh market as the repo market is gives basically two alternatives to the banks. Either they issue new securities and create new collateral or they sell the collateral which they can use only with high haircuts. This then leads to a price pressure on collateral, to even higher haircuts, etc. At some point liquidity and solvency of the market participants is triggered. But this destroys trust in the interbanking market to lend money - at some point there is a stand still or equivalently a break down if there is not a liquidity provider which can act differently than the banks trapped in the repo run channel: The central banks.

4.8.2 Bond and Stock Market

4.8.3 Swiss Bond and Stock Market

For equity we observe the volatility of market capitalization due to the events bust of dot.com bubble, 9/11, financial crisis 2008 and the boom period before this crisis. SMI comprises the most liquid titles of Swiss stocks. Adding smaller firms the SPI follows. Market capitalization of the bonds shows an almost linear growth. The picture becomes more detailed if one considers the number of products and issuers (Source: SIX-Group). We have for the last decade:

- The number of listed Swiss stocks and issuers is almost constant (about 70 stocks, 250 issuers).
- The number of foreign listed stocks fell from 172 in 2000 to 52 in 2010. The number of foreign issuers similarly dropped from 164 to 50 in the same period.
- For bonds, the number of domestic bonds almost halved (523 Bonds in 2010) and also the number of issuers fell by 40 percent to 117 issuers in 2010. In the foreign bond segment the number of bonds raised by 20 percent to a number of 943 in the year 2010.

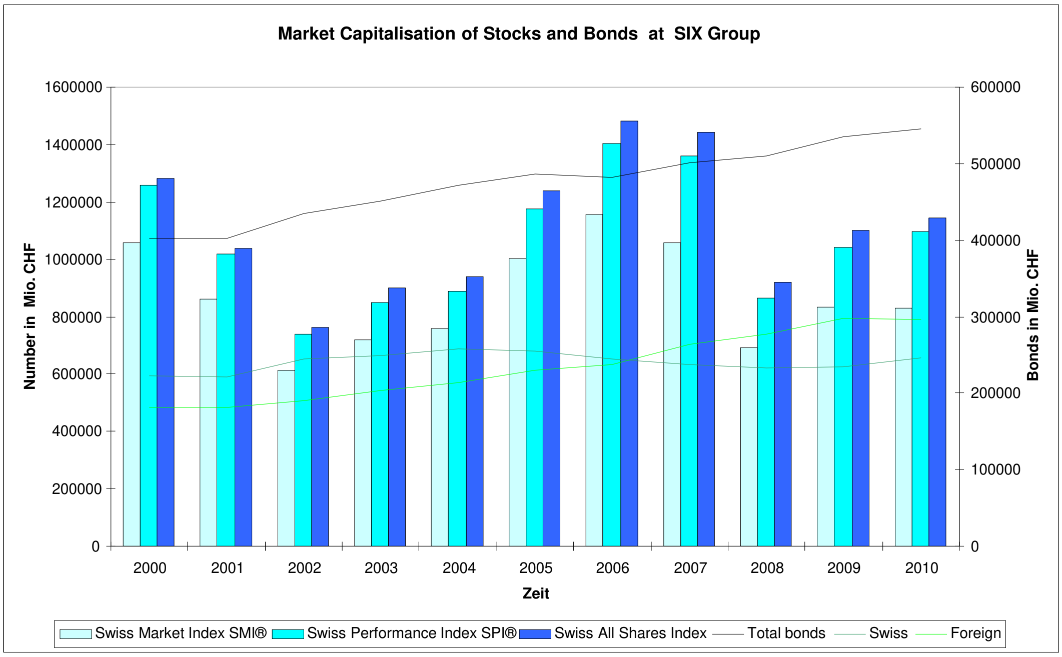


Figure 4.14: Market capitalization of stocks and bonds at Swiss Stock Exchange SIX-Group. 'Foreign' Bonds are bonds issued by foreign companies in CHF *Source: SIX-Group.*

Table 4.10 shows the evolution of the number of trades and volume for different products at SIX-Group.

	2001	2003	2005	2007	2008	2009	2010
Total Trades	11'456	13'927	17'954	35'339	45'186	34'772	34'978
Stocks	9'475	12'136	16'086	32'076	42'048	32'115	32'228
ETF, Funds	58	97	155	263	369	567	833
Bonds	783	674	531	461	559	592	521
Structured Products	1'139	1'020	1'185	2'539	2'211	1'498	1'396
Total Volume	868	693	927	1'780	1'653	924	967
Dom. Stocks	732	595	823	1'644	1'518	797	823
For. Stocks	22	10	18	12	5	3	3
ETF, Funds	4	7	14	26	34	42	60
Dom. CHF Bonds	31	20	12	14	21	19	18
For. CHF Bonds	46	42	34	24	25	28	26
Structured Products	33	18	27	61	51	34	37

Table 4.10: Trades and Volume at SIX-Group. Trades are in thousand; volume in Billion CHF. *Source:SIX-Group.*

The table shows the impressive growth of number of stock transactions which is mostly driven by Swiss stocks trading. Comparing this with the volume it follows that the average size of a trade significantly dropped. This observation, which is not only faced by SIX Group, and increasing demand for service functions are responsible for the pressure to lower the transaction costs. Finally, the table also shows the growth of ETF and the stand still of retail structured products (RSP).

Chapter 5

Retail Structured Products (RSP)

5.1 Definition and Structuring Idea

Retail Structured Products are in some sense an opposite investment vehicle to funds since most of them do not rely on the discretionary power of an asset manager but the final payoff is promised ex ante to the investor. This implies that banks have to generate with the initial investment amount the final payoff in any market circumstances: Trading, structuring, pricing and hedging are key disciplines for RSP.

Retail structured products offer investments into various asset classes, with many different payoff types and a fast time to market. These products represent contractual obligations from the issuer to the investor to pay out at maturity what is defined in the formulae on the term sheets. Contrary to funds, the obligation is related to the issuer non-defaulting, i.e. if the issuer defaults the (full) investment amount is lost (Lehman Brothers). Therefore, if a client buys a product from bank X which is only the distributor of the issuing bank Y it is the creditworthiness of Y which matters and not that one of X. For retail structured products the investment amounts are **not** treated as segregated capital. Retail structured products are much less regulated in some jurisdictions than investment funds. But this situation is changing. RSP are just called structured products in German speaking regions whereas structured products means in the anglo-saxon world products of structured finance, i.e. products which are derived from pooling assets in a special purpose vehicle and restructuring the pooled risks. This type of products is discussed in the Section Financial Crisis.

How are RSPs defined? The definition varies for different jurisdictions, sometimes a proper definition is missing but only a description exists. In the UK for example they differentiate between capital-at-risk and non-capital-at-risk products. In the former one conditional on the issuer non-defaulting the investor gets payed back a fixed amount of his initial investment at maturity as a minimum amount. A capital-at-risk product is defined as *... a product, other than a derivative, which provides an agreed level of income* *or growth over a specified investment period and displays the following characteristics:*

- *(a) the customer is exposed to a range of outcomes in respect of the return of initial capital invested;*
- *(b) the return of initial capital invested at the end of the investment period is linked by a pre-set formula to the performance of an index, a combination of indices, a 'basket' of selected stocks (typically from an index or indices), or other factor or combination of factors;*
- *(c) if the performance in (b) is within specified limits, repayment of initial capital invested occurs but if not, the customer could lose some or all of the initial capital invested.' Source: FSA Handbook.*

Point (c) defines that capital repayment is contingent on realization of events. A typical event are breaches of barriers by the underlying value.

Using this definition we consider **risk structuring** for an investor with the following preferences and views: He wishes a financial contract such that two investment motivations are jointly satisfied: Capital protection at maturity of the contract and participation in the performance of an underlying asset. More precisely, the investor's objectives are:

- Redemption of the investment capital at a minimum guaranteed percentage of the invested capital at maturity of the investment contract.
- Participation in the performance of an underlying asset.

Several issues arise in this context: First, how important are the two investment motivations for investors? Second, what means "capital guarantee" exactly? Third, how can the protection and participation seller make sure that she can fulfill the contractual obligations (hedging)? What is the fair price of such a contract (pricing)? How is "Participation" defined?

The main economic idea to structure a Capital-Guaranteed Product (CP) or a Structured Note is to exploit that the present value of say CHF 100 in 5 years is a lower amount today - the difference is used for participation. Then the seller of the guarantee, i.e. the issuer of the structured product, acts as follows:

- Suppose that in a given currency the annual interest rates are 2% where we do not specify what this rate is at, the moment. The PV of the guaranteed CHF 100 in 5 is
$$90 = (1 - 5 \times 0.02) \times 100 \ \ \text{CHF}$$
using linear compounding. If the issuer invests today CHF 90 in a zero-bond, then the capital guarantee promise in 5 years can be satisfied - if the issuer does not defaults.
- The amount of 10% is used to define participation for the investor.

Therefore, the investment product V_t consists of at least two products: A zero bond with price $p(t,T)$ at time t and maturity T and a participation product whose price depends on the price of the underlying asset S_t. In the simplest variant, the value of the product V_T at maturity T (the payoff profile or redemption amount) is determined as the product of the face value and the percentage change in the underlying asset's price during the term of the product:

$$V_T = N \times \left(1 + \max(0, b\frac{S_T - S_0}{S_0})\right)$$

with N the face value and b the participation rate, see Figure 5.5 for the payoff diagram. Rewriting, we get

$$V_T = N + \frac{bN}{S_0}\max(0, S_T - S_0) \ . \tag{5.1}$$

We note:

A '+'- sign in a payoff value formula in front of an financial instrument represents a long position and a '-'-sign a short position. The payoff formula (5.1) is written from a buyer's perspective.

Equation (5.1) shows that the payoff of the CP at maturity equals an investment in a zero bond and a long position in a European call option $C(S,K,T)$ with strike $K = S_0$. The number of options is equal to the face value divided by the initial price. Suppose that the call option is tradeable and liquid. Then, (5.1) is a replication of the payoff at maturity fixed in the contract. No arbitrage implies for the fair value of the contract V_0

$$V_0 = p(0,T) + \frac{bN}{S_0}C(0,S,K) \tag{5.2}$$

with $p(0,T)$ the zero bond and $C(S,K,0)$ the arbitrage free option price

$$C(0,S,K) = E^Q[D(0,T)\max(S_T - S_0, 0)]$$

under a risk neutral measure Q.

Consider an investor with different preferences:

- He beliefs that UBS stock is likely to raise over the next year.
- He believes that the stock will not raise strongly. He also prefers a partial capital protection if UBS stocks falls. He is in turn willing to give up the upside potential of the stock.
- He prefers a coupon which is larger than the UBS stock dividend.

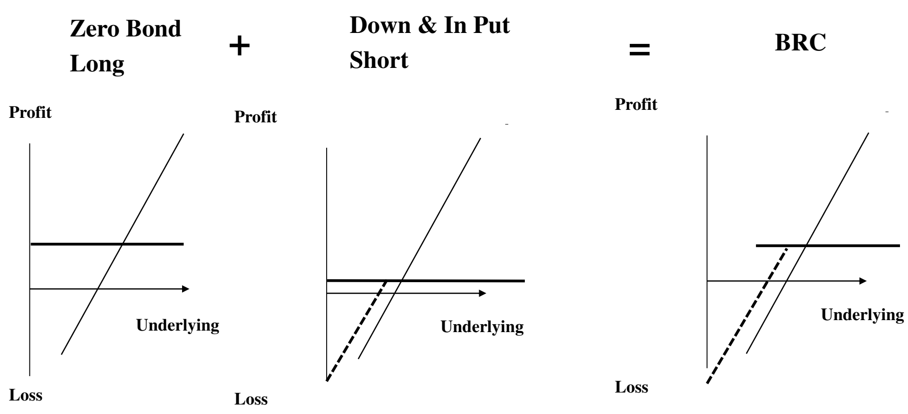


Figure 5.1: Structuring of the investor's preferences.

A structured product is able to match these investor preferences, see Figure 5.1.

The figure shows the payoff at maturity of a **Barrier Reverse Convertible (BRC)**. The investor gets **independent** of UBS stock price movements a coupon, say 10 percent. The investment amount is fully payed back **unless** the underlying value dropped below a barrier level during the life time of the product: The repayed capital amount is **contingent** on an event of a barrier hit. If the underlying value UBS once hits the barrier, the capital protection is knocked out and the payback at maturity is the UBS stock value at this date plus the 10 percent coupon. A BRC delivers a higher coupon than the stock dividend plus a contingent capital protection. The risk is the same as buying the UBS stock if risk is considered to be an adverse event which leads to a breach of the barrier. Why does anybody still buys stocks and not BRC? The investor gives up the stocks upside: The coupon is the maximum return.

How is it possible to pay a coupon which is larger than the stock's dividend and have also a partial capital protection? The answer is **structuring**. First, the investor gives up the upside potential and second, he sells an option. This makes a higher coupon value than the expected dividends possible. To understand this in more details, consider the replication of the BRC in Figure 5.1. The BRC payoff at maturity is replicated with two products:

- a long zero coupon bond position and

- a short down & in put (**DIP**). I.e. the investor sells a DIP - a barrier option on UBS. This money is used to generate the coupon value.

A **barrier put option** is characterized by a strike K and a barrier B. In our case $B < K$ - this leads to the expression 'down', see Figure 5.2. The payoff is: If UBS is for the whole time to maturity not breaching B, the option is worthless. Contrary if at any date the barrier is at least hit once the put option becomes active - the option is 'in'. In the BRC the investor is short the DIP - he sells it. If the barrier is hit a loss follows.

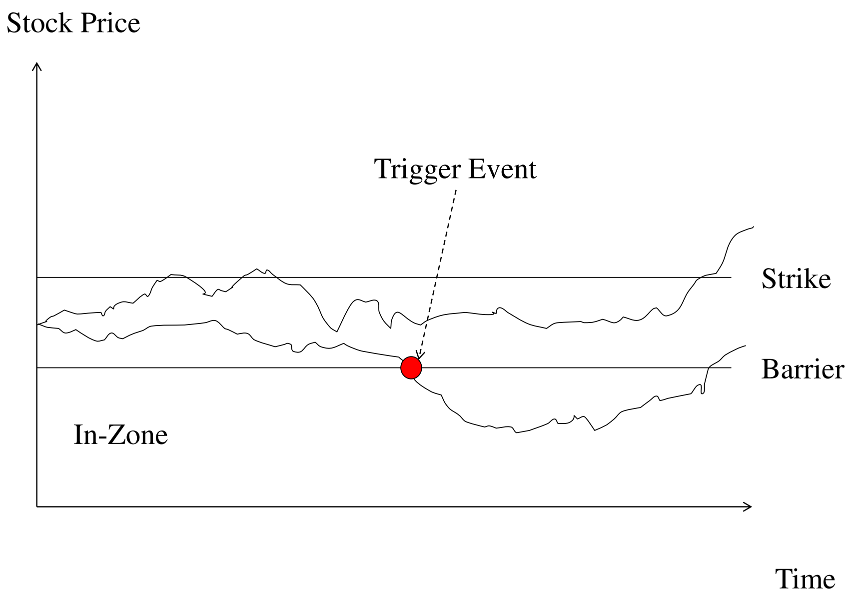


Figure 5.2: Down& In Put (DIP). The path which never hits the barrier never activates des DIP, i.e. the barrier option is worthless. The path which hits the barrier at least once gives to the barrier option a positive value.

Contrary to the problem to classify innovations in general, for RSP successful classification schemes exist. One of them is the Swiss Derivative Map from the Swiss Derivative Association. With minor adaption this map is also used in the European Structured Product Association. The Swiss map defines main categories:

- Capital protection.
- Yield Enhancement. BRC are a prominent product in this class.
- Participation, i.e. product which globally have linear payoff profile. 'Globally'

means that for some bounded region in the underlying value the payoff can be non-linear.

- Leverage Products. This is the class of warrants and mini futures.
- Reference Entity Products. In addition to the credit risk of the issuer, redemption is subject to the solvency (non-occurrence of a credit event) of the reference entity.

In each category there a sub categories, see Figure 5.3. The products are then represented in a single poster where for each product the payoff is graphically shown, the investment motivation and the risk and return characteristics are described.

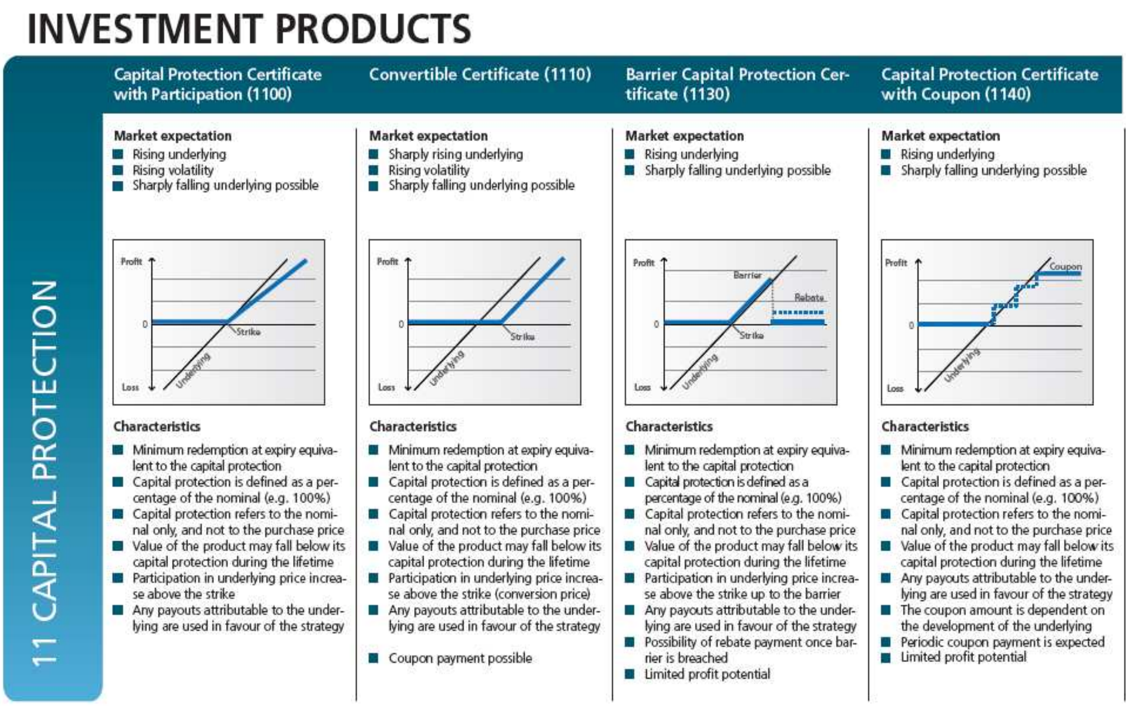


Figure 5.3: An excerpt of the Swiss Derivative Map for capital protection products. *Source: Swiss Structured Product Association, 2012.*

5.2 Structuring Capital Protection RSP

We reconsider the first example, i.e. equation (5.1) which shows that the payoff of the CP at maturity equals an investment in a zero bond and a long position in a European call option $C(S,K,T)$ with strike $K = S_0$. The number of options is equal to the face value divided by the initial price. Suppose that the call option is tradeable and liquid. Then, (5.1) is a replication of the payoff at maturity fixed in the contract. No arbitrage implies for the fair value of the contract V_0

$$V_0 = p(0,T) + \frac{bN}{S_0} C(0,S,K) \tag{5.3}$$

with $p(0,T)$ the zero bond and $C(S,K,0)$ the arbitrage free option price. Not all structured products can be decomposed into a sum of basic products - stocks, forwards, bonds and vanilla options. If such a decomposition is not possible, different taxation rules can apply. In Switzerland such non-decomposable products are classified as not-transparent and taxed different than transparent products.

Figure 5.4 shows the evolution of a 5 year CP in two extreme scenarios: Once the underlying value is simulated with a strong negative drift and in the other scenario the drift is strongly positive. The negative scenario shows that value of capital protection compared to a long position in the underlying. In the positive scenario the price of capital protection follows: Participation is capped to 120 percent and the investor gives up any further upside.

There are many variation of the so far discussed CP product. Figure 5.5 shows the case with a variable capital protection level a and participation rate b. The CP payoff reads:

$$V_T = aN + \frac{bN}{S_0} \max(0, S_T - S_0)\ . \tag{5.4}$$

Example

Consider a $5y$, $a = 100\%$-level CP product with underlying value the Swiss Market Index (SMI) of Swiss Blue Chips. The product is issued October, 28 2005. The closing value of SMI at issuance date is 6'874 points which is also the strike value of the call option in the replication. The face value is CHF 5'000. Using Black and Scholes with a volatility of 18%, risk free rate of 3 percent leads to a participation of $b = 58\%$. To see this, we note that the Black and Scholes Price of the option is

$$C(0) = S_0\Phi(d_1) - Ke^{-RT}\Phi(d_2)$$

with

$$R = 1.03\ ,\ d_1 = \frac{\ln(S_0/K) + RT}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2}\ ,\ d_2 = d_1 - \sigma\sqrt{T} = \frac{\ln(S_0/K) + RT}{\sigma\sqrt{T}} - \frac{\sigma\sqrt{T}}{2}\ .$$

In our example $d_1 = 0.60981, d_2 = 0.20732$ and $\Phi(d_1) = 0.72901, \Phi(d_2) = 0.58212$. This implies for the option the price $C_0 = 1'640$. Solving

$$5'000 = \frac{5'000}{(1+0.03)^5} + \frac{b \times 5'000}{6874} 1'640$$

implies the participation rate. Replication from the buyer's view gives the following redemption at maturity:

$$V_T = 1 \times 5'000 + \frac{0.58 \times 5'000}{6'874} \max(0, S_T - 6'874)\ . \tag{5.5}$$

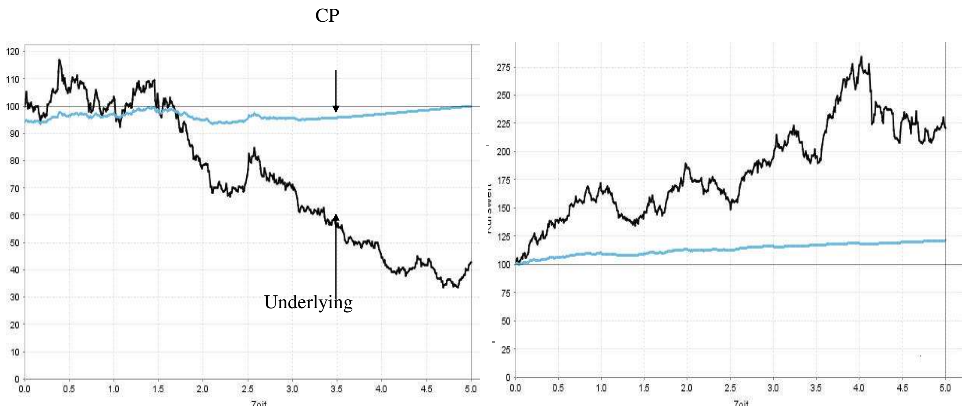


Figure 5.4: Capped CP with capital protection 100 percent, participation 100 percent and cap level 120 percent. The figure shows that the CP may well value below the capital protection level before maturity. Left Panel: Drift of the underlying value of −5 percent p.a. Right Panel: Drift of the underlying value of +5 percent p.a.*Source: Zurich Cantonal Bank. Dynamic Derivatives Education Tool.*

The opposite position of the market makers is as follows. Suppose that a total amount of CHF $100'000$ is invested in the product, i.e. the market maker is short a notional amount of $100'000$ zero bonds with $5y$ maturity. The difference between the notional amount and the present value of the bonds is used to sell call options, i.e. the market maker is short

$$8.4367 = \frac{0.58 \times 5'000}{6'874}\frac{100'000}{5'000}$$

$5y$ calls on the SMI. Conditional on the terminal SMI-value, different return calculations follow. If SMI is up by 10% at maturity, i.e. $S_T = 1.1 \times 6'874 = 7'562$, redemption equals

$$\begin{aligned} V_T &= 1 \times 5'000 + \frac{0.58 \times 5'000}{6'874}\max(0, 7'562 - 6'874) \\ &= 5'000 + \frac{0.58 \times 5'000}{6'874}(7'562 - 6'874) = 5'290 \ . \end{aligned}$$

This is equivalent to a 5.80% five-year return of the invested amount. Contrary, assume that SMI is down by 10%. Then the call is out-of-the-money, i.e. worthless. Redemption

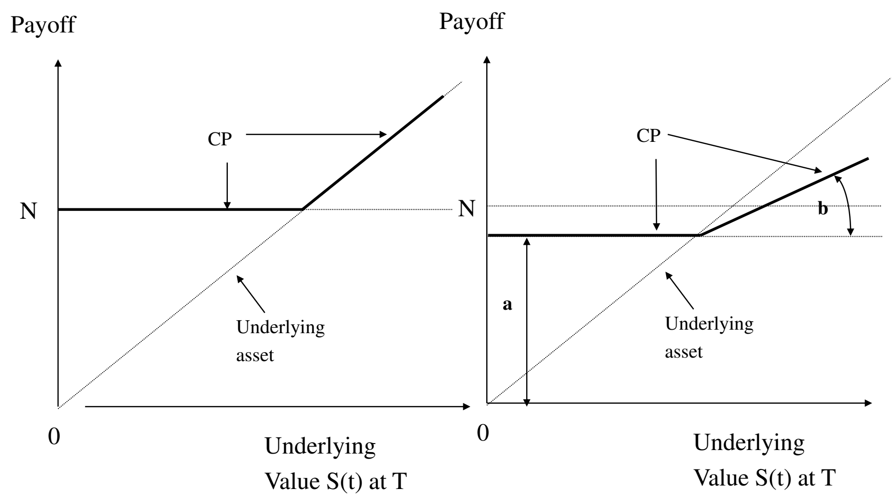


Figure 5.5: Payoffs for CP products. Left panel: CP product with 100% capital guarantee and 100% participation. Right panel: CP product with a% capital guarantee and b% participation rate.

then equals the capital protected amount of 5′000. Many people speak about a opportunity loss in this scenario since the investor could have invested the CHF 5′000 in bond with the same credit risk as the CP product issuer with a corresponding interest rate return. This is an odd argument since it compares the performance of two products for two different types of investors: An investor which seeks both capital protection and participation in a stock market compared to an investor which prefers to participate in an interest rate product. Such a comparison makes little sense.

5.2.1 Variations of the Payoff

If interest rates are low, only a few is left to buy the call option and to participate. A low participation rate follows. This rate can be increased if the capital protection level a is lowered. In an environment with low interest rates the above structures are hard to sell since participation is too low and/or people require full capital protection. A possible solution in this case is to give up the high return region, i.e. to cap the payoff as shown in Figure 5.6.

The investors give up the gain surface B in order to finance the loss protection zone

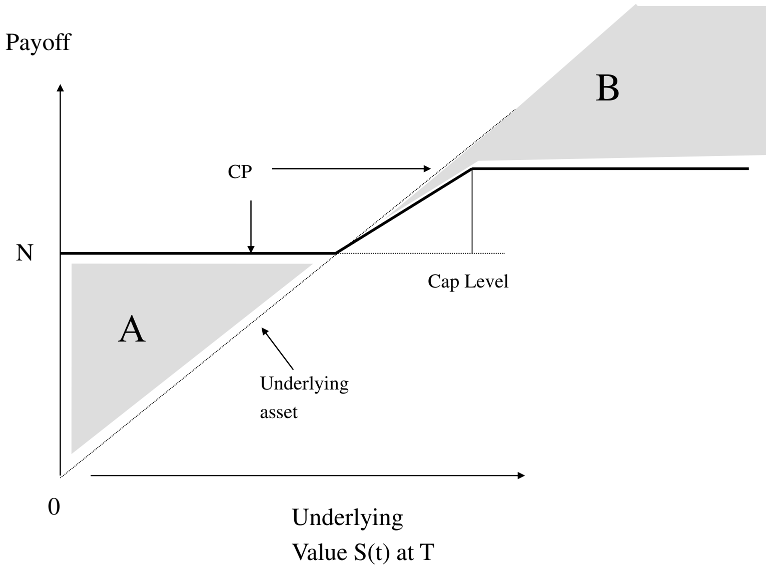


Figure 5.6: Capped payoff for CP product.

A. We get the desired payoff if the investor is long a call with strike K_1 and short a call with a higher strike K_2, see Figure 5.6. Therefore, the investor is long a zero bond, long a call with strike K_1 and short a call with strike K_2 where $K_1 < K_2$. This long-short strategy is a call-spread strategy.

Example

Pricing of a capped CP using Black and Scholes for different parameters, see Table 5.1. The table shows in particular how an increasing volatility makes the options more expansive.

Price of	$\sigma = 20\%, r = 2\%$	$\sigma = 30\%, r = 2\%$	$\sigma = 20\%, r = 3\%$
Zero bond	81.52%	81.52%	77.36%
All coupons	9.42%	9.42%	9.15%
Long Call option	22.02%	30.04%	24.33%
Short Call option	-12.96%	-20.97%	-11.11%
CP	100%	100%	100%

Table 5.1: Price of a capped CP with capital protection 90 percent, 2 percent annual coupon payments, 5 years maturity and 100 percent participation.

Consider an investor which not only wishes full capital protection, $a = 1$, but also requires a minimum guaranteed payoff or return. This minimum return f is called the floor or floor rate. The payoff of the CP reads

$$V_T = N\left(1+\max\left(f, b\frac{I_T - I_0}{I_0}\right)\right) =: N + O_T \tag{5.6}$$

for an index I as underlying value. O_T is the option payoff which is equivalent to

$$O_T = N\left(f+\max\left(0, b\frac{I_T - I_0}{I_0} - f\right)\right) \tag{5.7}$$

where we used $\max(a,b) = a + \max(0, b-a)$. Using $\max(0, ab+c) = a\max(0, b+c/a)$ if $a > 0$, we get

$$O_T = N\left(f+\frac{bN}{I_0}\max\left(0, I_T - I_0(1+\frac{f}{b})\right)\right) . \tag{5.8}$$

The payment is equal to a fixed amount plus an index-linked amount. The latter one is the same as the payoff of b/I_0 vanilla calls with strike $I_0(1+\frac{f}{b})$. This strike is higher than in the case with floor zero. The floor raises the option price or lowers the participation rate if the capital protection level is kept fixed. Using the put-call parity, the payoff is equivalent to

$$V_T = \frac{bN}{I_0}(I_T - I_0) + \frac{bN}{I_0}\max\left(0, I_0(1+\frac{f}{b}) - I_T\right) . \tag{5.9}$$

The issuer can choose the valuation formula which is more suitable, i.e. the replication with the more liquid options for example.

So far the issue price of a structured products was set equal to the sum of fair prices of the replication portfolio. But there are cost components. In practice, there is a margin. The investor pays 100 and the replication portfolio or product has a value of say 98. The amount of the margin, 2 units in the example above, depends on several aspects. First, for mass products margins are generally lower than for a tailor-made products or for a product with a complex payoff. In the latter case trading will ask for buffer against hedge risks. Also the more exotic the underlying value is the higher is the margin. Besides a net profit, the margin has to cover the production costs (trading, sales, quants, mid-office, back-office, risk function, compliance), has to cover the capital-at-risk costs (the costs for the risk capital needed to run the business) and marketing costs. We consider this in the Section Dark Side of Innovation in more details.

5.2.2 Increasing Participation

There are several ways to increase the participation rate in CPs. We follow in some parts closely Kat (2008).

5.2.2.1 Higher Base and Longer Maturity

We allow for a different base rate g which is not necessarily equal to one. The option part of the payoff

$$V_T = N\left(1 + \max\left(f, b\frac{I_T - gI_0}{I_0}\right)\right) \quad (5.10)$$

is equivalent to

$$O_T = N\left(f + \frac{bN}{I_0}\max\left(0, I_T - I_0(g + \frac{f}{b})\right)\right) . \quad (5.11)$$

This shows that the call has a higher strike for $g > 1$ than for $g = 1$. The call becomes cheaper with an increasing base rate g. To obtain the same value V_T as for the case $g = 1$, one has to increase the participation rate b. If an investor wishes to participate at a higher degree he has to accept a higher base rate: The index level has to raise to a higher level before participation starts.

If an investor chooses a longer maturity two effects follow. First, there is more money left to invest into the option since the present value of bond for capital protection decreases. Second, the options are more costly if maturity increases if one assumes a flat volatility term structure. The longer maturity is, the further away can the underlying value deviate from strike value. This increases the expected option payoff both for calls and puts since the investor. Furthermore, the martingale property of the discounted underlying value sets its drift equal to the risk free rate: The longer maturity is the more is a call in the money. This results in a strong growth of call option premium with increasing maturity. For put options the option premium drops quickly close to zero. But the term structure of volatility is not flat. If long-dated volatility is higher than short dated volatility then option prices will increase faster. If the volatility term structure is steep enough, the net effect on the bond and option part may favor shorter maturities. This holds for uncapped participation. If participation is capped, the opposite holds since the bank is writing options.

Summarizing, an increase of maturity date has ambiguous effects on the participation rate. While the time value of the capital protection leads to an increase for increasing maturities, the value of the options depends on the volatility structure and its evolution over time.

5.2.2.2 General Piecewise Payoff

Once an investment idea is setup, it becomes important in to fine tune the payoff. A linear payoff is for example replaced by a piecewise linear one such that in different payoff regions different participation rates follow. This leads to the following exercise: Construct a piecewise linear payoff with kinks $0 < K_1 < K_2 < K_3$ and participation

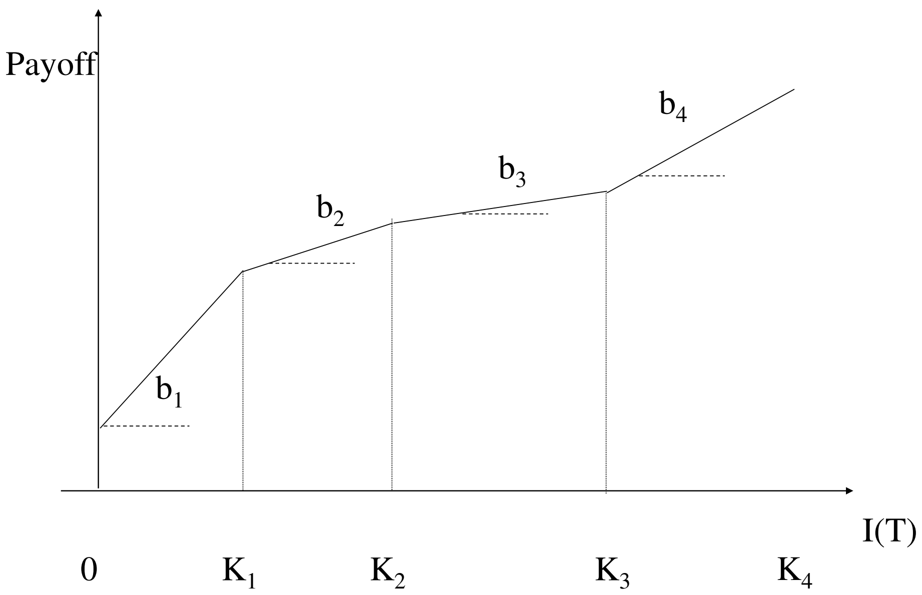


Figure 5.7: Constructing a piecewise linear payoff.

rates b_1, b_2, b_3 and b_3 where the rate applies in the intervals b_1 in $[0, K_1]$, b_2 in $(K_1, K_2]$, b_3 in $(K_2, K_3]$ and b_4 in (K_3, ∞), see Figure 5.7.

The piecewise linear payoff at maturity then reads

$$\begin{aligned} V_T &= b_1 \min(I_T, K_1) + b_2 \min(\max(0, I_T - K_1), K_2) \\ &+ b_3 \min(\max(0, I_T - K_2), K_3) + b_4 \max(0, I_T - K_3) \ . \end{aligned} \tag{5.12}$$

To make this plausible, for $I_T < K_1$ we get $b_1 I_T$. For $K_1 < I_T < K_2$ follows

$$b_1 K_1 + b_2(I_T - K_1) \ .$$

For $K_2 < I_T < K_3$

$$b_1 K_1 + b_2 K_2 + b_3(I_T - K_2) \ .$$

and finally $K_3 < I_T$

$$b_1 K_1 + b_2 K_2 + b_3 K_3 + b_4(I_T - K_3) \ .$$

This is a plausibility check for the formula (5.12). This formula is readily generalized to the case of n kinks.

As an application, consider the piecewise segmentation of the following capped CP:

$$V_T = N(1 + \min(c, \max(f, b\frac{I_T - I_0}{I_0})) = N + O_T$$

with $c > 0$ the cap value. The option payoff can be rewritten as

$$O_T = fN + b\frac{N}{I_0}\left(\max(0, I_T - I_0(1+f/b)) - \max(0, I_T - I_0(1+c/b))\right) \ .$$

This is a call spread. Again, an increase in the floor rate f lowers participation and an increasing cap rate c makes the option cheaper - since it is short position participation then also drops. This shows the rationale for capped capital protection: Selling an option the investor restricts the upside potential which in turn leaves more money for participation.

5.2.3 Barrier Options

One can increase the participation rate by replacing the call option by a up& out call barrier option: If the underlying instrument never reaches a barrier value, which is above strike, the investor participates. If the barrier is at least once touched, participation becomes worth-less and the capital protected amount is repaid or a rebate value follows. Such an option is an up-and-out call option which is cheaper than a vanilla call. Combining such options with capital protection leads to Shark Notes SRP, see Figure 5.8 for a replication. A Shark Note is replicated with long positions in a zero bond, an up& out

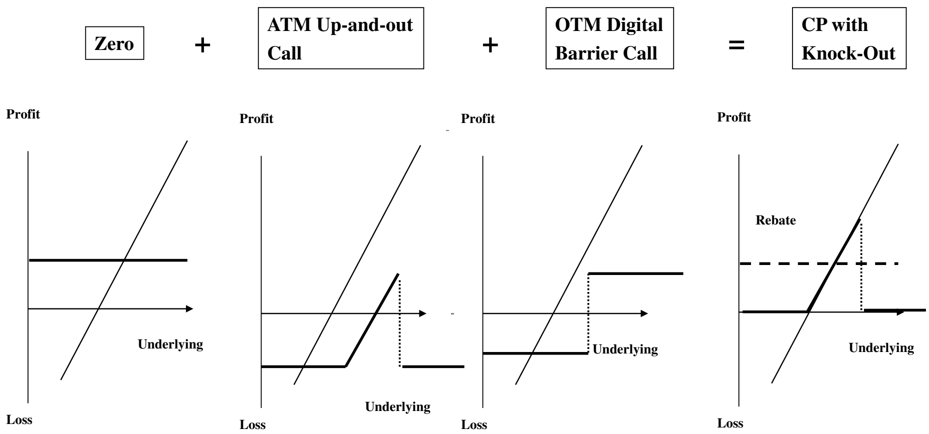


Figure 5.8: Replication of a Shark Note.

ATM call barrier option and an OTM digital barrier call. This is not a vanilla replication. A specification of such a contract is given in Table 5.2. Table 5.3 summarizes some

Maturity	1 year
Issue Price	100 percent
Underlying	DAX
Reference Value Underlying	6'100 points
Strike	100 percent
Knock-out-Level	140 percent
Participation Rate	100 percent
Rebate	5 percent
Capital Guarantee	100 percent

Table 5.2: Specification of a Shark Note contract.

sensitivities of the Shark Note.

Factor	Zero	Up-and-out Call	Digital	Shark Note
Underlying up	-	ambiguous	up	mostly up
Strike up	-	down	-	down
Knock-out level up	-	up	down	mostly up
Time to maturity increasing	down	ambiguous	ambiguous	ambiguous
Implied volatility up	-	ambiguous	up	ambiguous

Table 5.3: Sensitivities of a Shark Note contract. In the first column are the risk factors. The remaining columns indicate the impact if a risk factor is shocked in the indicated direction.

We consider **barrier options** in more details. Barrier options are the oldest as well as the most successful type of all exotic options. They were traded in the US OTC market before the Black and Scholes pricing method was known since 1967. Barrier options are exotic options since their payoff is path-dependent, i.e. the payoff at a given date depend on whether or not the underlying asset has reached some barrier price during the life of the option. Barrier options can have three possible features. First, the position of the barrier relative to the strike (up if $B > K$ or down else). Second, whether the option becomes worthless once the barrier is touched (out) or whether the option has a value (in). Third, whether the option is OTM once a barrier is reached (regular) or ITM (reverse). The path dependency gives European barrier options a flavor of an American option. Since the critical exercise boundary - the barrier - is known in advance, they are simpler to price than pure American options. Rubinstein and Reiner (1991) derived a closed-form analytical solution following the dynamic hedging approach of Black and Scholes. This pricing formula can in principle be used to hedge barrier options continuously. But the dynamic Delta-hedging in a Black-Scholes framework leads to difficulties. The Delta of the barrier options is very sensitive to changes in the price of the underlying close to the barrier. Hence, traders need to rebalance the hedge very often. This leads to large transaction costs and is a challenge for the operations of large barrier option portfolios. Figure 5.9 compares the a down-and-in put barrier (DIP) and a vanilla option and shows the Delta and Vega of the barrier option. The plot shows that near the barrier

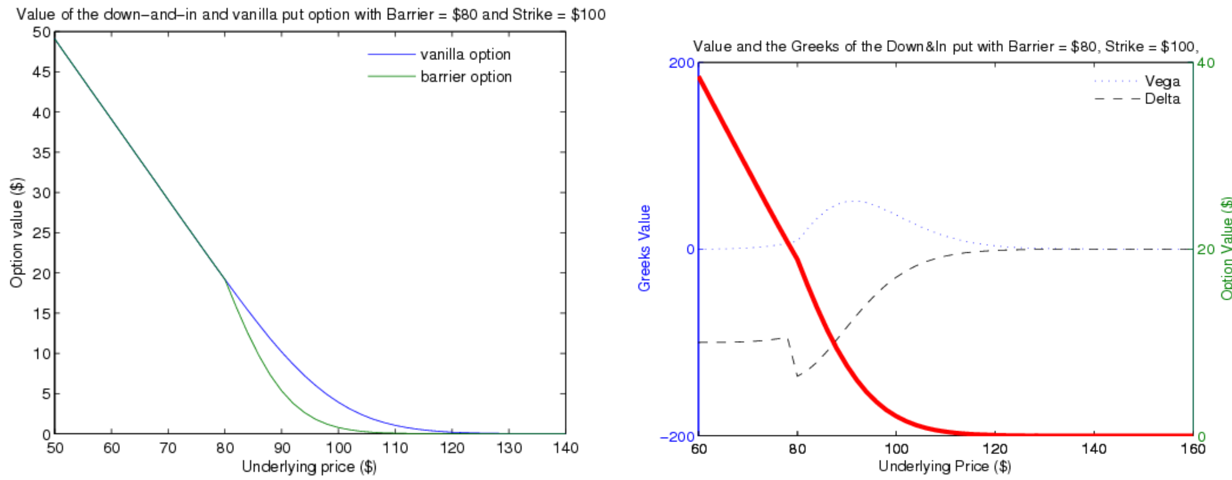


Figure 5.9: DIP and vanilla put. *Source: N. Dolgova, MAS Finance Univ. of Zurich and ETH Zurich, 2008*

the Delta becomes very sensitive - small changes in the underlying lead to large changes in the Delta. If furthermore the option is close to maturity the Delta in the region where the underlying is close to the barrier becomes a discontinuous function. This makes continuous time hedging difficult. Alternative hedging approaches were developed. We consider them in the Barrier Option Section.

A barrier option is cheaper than the vanilla counterpart since a barrier options activates the payoff contingent on an event. This is true in general and is a major rationale to buy barrier options.

We consider some **facts** about barrier options. Combining a knock-in barrier option and a knock-out barrier option with the same characteristics produces an European plain vanilla option. This is because as soon as the barrier is reached, the knock-out options disappears, but the knock-in option immediately appears. A similar arguments holds if the barrier is not reached.

It is in general more complicated to hedge barrier options than their corresponding vanilla options and hedging results are normally less intuitive than for vanilla options. An exception are barrier options with **strike equal to the barrier**, see below. Changes in implied volatility affect plain vanilla options in a uniform manner: an increasing volatility increases the option price. With barrier options, increasing volatility can affect the premium in **two opposing ways**. Increasing volatility increases the probability

- that the option will be in-the-money and therefore increases its value;
- that the barrier will be reached, thus decreasing the value of a knock-out option and increasing the value of a knock-in option.

Therefore, an increase in implied volatility has a double effect on a knock-in option and two opposing effects on a knocked-out option.

There are many variations of barrier options:[1]

- Daily versus continuous **monitoring** of the barrier. A knock-out option where the barrier is monitored only daily will be more expensive than the equivalent continuously monitored barrier, since an observation with the barrier being hit is less likely to occur. Typically the difference in price will be around 5 percent of the option premium.
- **Multiple barriers**: There can be more than one barrier on any option, and one barrier can be contingent on another barrier being breached (or not). A 'double-out' option with barriers above and below spot will be knocked out if either of the barriers is hit. An ATM call knocking out at 110 percent and 95 percent will be cheaper than a ATM call knocking-out at 110 percent.
- **Early exercise**: Barrier options are typically traded as European with no early exercise possible. An early exercise feature will make knock-out options more expensive.
- **Resetting barriers**: This is a barrier option which turns into another barrier option with a different barrier level when the initial barrier is hit. It can be constructed from sequences of regular barriers with rebates.
- **Forward starting and early-ending barrier options**: The barrier may only exist for part of the lifetime of the options.

[1] The list is an abbreviated form of JPMorgan, European Equity Derivatives Stratgey, 2007.

- **Soft barriers**: A soft barrier allows a contract to be gradually knocked in or out. Such a contract specifies an upper or lower barrier level and knocks the option in or out proportional to the point reached in the range. For example an up-and-out call with strike 100 and soft barrier 110-120 will have half its intrinsic value knocked out if the maximum level reached over the lifetime is 115 and will be completely knocked out if the underlying ever reaches 120.

- **Rebates** could be paid as a form of compensation, if a knock-out barrier is reached. This feature increases the premium of the barrier option.

- **Ratchet or Cliquet Options**. Suppose an investor buys an ATM index put to protect a portfolio. If the index rallies the investor wants to roll up the protection by selling the original put and buying a new ATM put. However, at this time the new ATM put will naturally be much more expensive than the existing, now OTM put. The investor could pre-position himself for this situation by trading in barrier options. For example suppose the investor bought an ATM up-and-out put with knockout at 105 and a 105-strike put knocking in at 105. This structure will then automatically 'ratchet' the protection up to an ATM put as soon as the underlying reaches 105. This sequence can be further extended to lock in protection at regular intervals. Such structures provide downside protection with a significant upside potential. Since this options lead to resetting the strikes **forward starting options** matter. A forward starting call is defined in a two period model

$$C_{\texttt{FwS}}(T) = \max(\frac{I_2}{I_1} - K, 0) \tag{5.13}$$

as opposed to the ordinary call

$$C(T) = \max(\frac{I_2}{I_0} - K, 0) \ .$$

If we take out I_1 of the maximum it follows that in the forward start case the strike becomes time dependent whereas in the ordinary call case the strike is fixed and constant at time 0. The investor therefore pays at 0 for an option with a strike known only at time 1. Hence, forward start options are used for investors which want exposure to an underlying asset starting at a future date but they prefer to pay or receive an amount today at 0. To price such options one needs to use the volatility at a future date. This makes the use of the Black and Scholes model cumbersome since it provides no model of how the present volatility structure evolves over time. Local volatility models, which are commonly used for barrier options, also turn out to be not adequate but instead a SABR model is used Given the forward start calls and puts the cliquet options are simply floored and capped variations of these options. The mechanism can be quite complicated as we have seen in the pension fund example. Cliquet are typically defined on indices and the usual variations such as the worst-off structure, reverse cliquet applies. Cliquet option payoff's are of the form

$$V_{\texttt{cliquet}}(T) = \min\left(C_{\texttt{glob}}, \max\left(F_{\texttt{glob}}, \sum_{j=1}^{M} \min\left(C_j, \max\left(b\left(\frac{I_j}{I_{j-1}} - K_j\right), F_j\right)\right)\right)\right) \tag{5.14}$$

with $F_{\texttt{glob}}$ the global floor, $C_{\texttt{glob}}$ the global cap, C_j and F_j the local cap and floor respectively and M the number of legs. If we consider

$$V_{\texttt{cliquet}}(T) = \min\left(\infty, \max\left(0, \sum_{j=1}^{2} \min\left(5\%, \max\left(b\left(\frac{I_j}{I_{j-1}} - 100\right), 0\right)\right)\right)\right)$$

the price of the options at time zero in a 2y product with 100 percent participation is given by:

$$\begin{aligned}
V_{\texttt{cliquet}}(0) &= e^{-rT}E^Q\left(V_{\texttt{cliquet}}(T)\right) \\
&= e^{-rT}E^Q\left(\min\left(5\%, \max\left(\frac{I_1}{I_0} - 100, 0\right)\right)\right) \\
&+ e^{-rT}E^Q\left(\min\left(5\%, \max\left(\frac{I_2}{I_1} - 100, 0\right)\right)\right) \\
&= e^{-rT}E^Q\left(\max\left(\frac{I_1}{I_0} - 100, 0\right) - \max\left(\frac{I_1}{I_0} - 105, 0\right)\right) \\
&+ e^{-rT}E^Q\left(\max\left(\frac{I_2}{I_1} - 100, 0\right) - \max\left(\frac{I_2}{I_1} - 105, 0\right)\right) ,
\end{aligned}$$

i.e. the cliquet option can be decomposed into two forward starting call spreads with deferred settlement.

- **Ladders**. Closely related to the ratchet structure is a ladder. A ladder option is a call or put option on a index, which periodically resets when the underlying trades trough specified trigger levels called "rungs", at the same time, knocking in the profit between the old and new strike. At maturity, it pays the maximum of the index and the underlying itself minus the strike floored to zero for the call version. The name comes from the fact that the trigger strikes play the same role as rungs on a ladder. The ladder option can be structured to have its strikes reset in either one or both directions, allowing a great flexibility in the payoff. For example, suppose we require a structure which locks in profits each time the underlying increases 10 points – starting from 100. We can buy a sequence of knock-out calls with strike and barrier 10 points apart, each paying a rebate of 10, ending with a vanilla call. E.g. an ATM call with KO (knock-out) at 110 with rebate of 10, a 110-call with KO at 120, with a rebate of 10 and a vanilla 120 call. Once the underlying reaches 110, the first call knocks out with rebate of 10 – locking in the profit up to this point, and the 110-call is now in the money, continuing the upside exposure.

- In the limit with the rungs arbitrarily close together the ladder structure becomes a **lookback call option** – paying the maximum value reached by the underlying over the lifetime of the option. A lookback call typically costs twice the price of the corresponding vanilla call – and places an upper bound on the cost of a ladder. Typically with rungs around 10% apart a ladder would cost around 1.5 times the equivalent European.

- **Range notes**. A range note is simple structure paying a coupon each day whilst an underlying remains in a range and ceasing when it moves outside. It is just a one-touch double American digital option. A similar structure is a range-accrual which pays a fixed amount for each day an underlying remains in the range, but pays zero outside the range. The difference is that the range-accrual will continue paying when the underlying re-enters the range. Such a structure is constructed as a strip of double European digitals, one maturing each day.

We consider the hedging of simple down-and-out calls - i.e. $B = K$ - using the underlying. Without loss of generality $B < S_0$. This option has no optionality or time-value. To see this, suppose we sold the knock-out call and that we wish to hedge. We simply charge the intrinsic value as a premium $S_0 - K$ and hedge by buying one unit of the underlying. Then:

- The barrier is never hit. In this case the call ends in-the-money, costing us $S_T - K$. We make $S_T - S_0$ from holding the underlying and retain the $S_0 - K$ premium charged. Total profit and loss is zero, i.e.

$$-S_T + K + S_T - S_0 + S_0 - K = 0 \ .$$

- If the barrier is hit we liquidate the position in the underlying. We realize a loss of $S_0 - B$. Since $B = K$, this equals the premium we charged for the knock-out and total profit and loss is zero.

In theory the hedge is perfect. This option hedge is simpler than for the corresponding vanilla option since the barrier/strike can only be crossed once. This allows us to completely liquidate our position at the barrier without worrying about having to buy it back if it re-crosses – and hence avoiding further hedging. We have assumed that as soon as the underlying crosses the barrier we can sell our position. But in reality if the underlying gaps through the barrier we will make a loss. To compensate for this we should increase the premium charged by this expected slippage multiplied by the probability of hitting the barrier. This defines **gap risk.**

5.2.4 Changing the Reference Index: Dividend Yields and Correlation

An index can be changed for several reasons: Higher dividend yields, introducing correlation to raise participation or changing to foreign indices. Dividends lower the upward trend of stocks since the stock price drops at the ex-dividend dates. Therefore higher dividend yields make calls cheaper and puts more expansive. To raise the participation rate one switches to indices or stocks which pay higher dividend yields.

We derive the price of a dividend paying call. Suppose that the stock pays dividends at continuous rate d. This is similar to the interest rate effect, i.e. 1 CHF grows with $e^{r(T-t)}$ under continuous compounding and 1 share grows with $e^{d(T-t)}$. Therefore, the fair price of a call is not based on 1 share but on $e^{-d(T-t)}S$-shares - else the Black and Scholes formula remains the same. Thus the change

$$S \to e^{-d\tau}S$$

in the Black and Scholes formula provides us with the correct option price for a dividend paying stock. If dividend are paid the underlying value jumps. But since dividends goes to the holder of the stock not to the holder of an option on the stock the value of the option is continuous over the dividend paying date. The option value shortly before and after dividend payments is the same. If t_- and t_+ are the two dates close to the dividend date this means

$$C(S, t_-) = C(S - D, t_+)$$

for a dividend value D. Although the value of an option is invariant, the Delta changes since the a dividend payment defines a new time value curve for a call option, see Figure 5.10.

Using baskets instead of single stock underlyings one introduces **correlation** as a new parameter and one alters the volatility. The volatility of an index is in general lower than the volatility of the constituents. This is due to the imperfect correlation of the basket constituents. But lower volatility means cheaper call options which leads to a higher participation rate. To raise participation one searches for low correlated underlyings or even negatively correlated ones. Although a significant increase in the participation rate is possible, in practice the effects are small. This has two reasons. First, the search for uncorrelated constituents makes the prices goes up. That is derivative firms raises the prices for baskets with a low correlation. Second, correlation risk is not straightforward to deal with. This means that only highly professional derivative firms offer options on baskets and in particular on less liquid ones. But his makes the supply side stronger which leads to higher pricings.

5.2.5 Changing the Reference Index: Foreign Reference Index

No arbitrage implies that the drift of the risky asset is equal to the risk free rate. Exchanges rates have a trend upwards or downwards at a rate which is equal to the interest rate differential. If we replace the domestic index by a foreign index, which is denominated in a currency with a lower interest rate than the domestic one, we can raise the participation rate. A possible design is the payoff

$$V_T = N\left(1 + \frac{E_{T,\mathtt{D-F}}}{E_{0,\mathtt{D-F}}} \max(f, b\frac{I_T - I_0}{I_0})\right) .$$

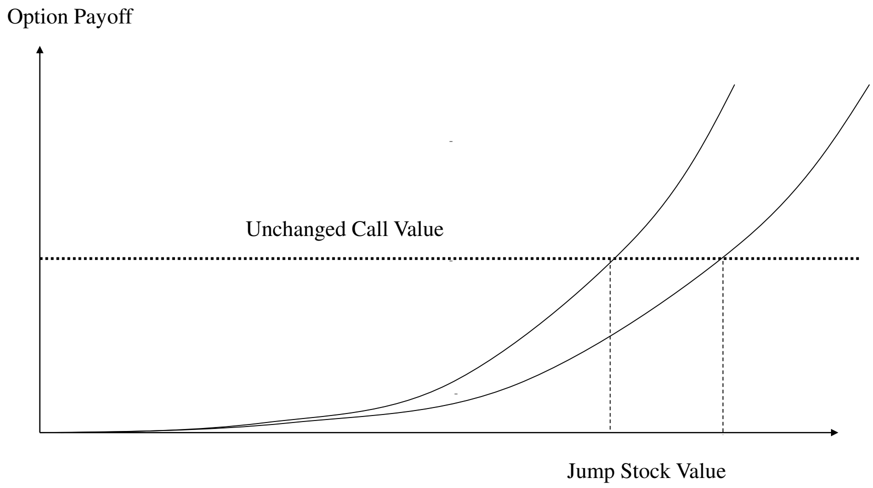


Figure 5.10: Impact of dividends on option prices.

The index I is measured in the foreign currency, the floor f floors a foreign currency amount, $E_{t,\mathtt{D-F}}$ is the spot exchange rate in units of DOM (domestic currency) per unit of FOR (foreign currency) and $\frac{E_{T,\mathtt{D-F}}}{E_{0,\mathtt{D-F}}}$ converts the foreign amount at maturity into a domestic one. Investors are in this case exposed to exchange rate risk due to he conversion factor. To check the dimensions let I be in EUR and the exchange EURCHF:

$$V_T \simeq \text{EUR}\left(1 + \frac{\text{EURCHF}}{\text{EUR/HF}}\max(\%)\right) \simeq \text{EUR} .$$

Separating the option part in the above payoff, the budget equation at time 0 implies the participation rate:

$$PV(O_T) = e^{-Tr_{\mathtt{D}}}\, fN + bNf/I_0 \times C_{\mathrm{F}}\,(0, I_0(1+f/b), T)$$

with $C_{\mathrm{F}}\,(0, I_0(1+f/b), T)$ a foreign currency call. We assume in the following discussion that all parameters except those under discussion are the same in either currency. If the foreign and domestic interest rate agree, the same participation rate follows. If foreign rates are lower than domestic ones, an increasing participation is obtained. The equation determining the participation rate is solved numerically. We know that already small differences in the rates lead to substantial differences in participation rates since the call is sensitive to the interest rate.

Before we continue the discussion it is worth to introduce to some concepts of the **FX asset class.**

Definition 5.2.1. *The FX rate is the price of one currency in terms of another currency. Two currencies make a pair. For example, if we take Euro and USD as an exchange rate, the default quotation is EURUSD. EUR is the base currency and USD the numeraire one. FX is the price of the base currency EUR in terms of USD. Equivalent is to call the numeraire currency the domestic currency (DOM) and the other one the foreign currency (FOR).*

FX rates are expressed as five-digit numbers, say 1.2245 for CHFEUR. The fifth digit (5) is called a pip, 100 pips make a figure. A spot contract is a bilateral contract to exchange the base currency amounts against an amount of the numeraire currency equal to the spot FX rate. An European plain vanilla FX option **USD call CHF put** is a spot contract where the buyer of the contract has the right to enter at expiry into a spot contract to sell (buy) the notional amount of USD (CHF) at the strike FX rate level K. The notional amount N in the USD base currency is exchanged against NK units of the numeraire currency. The buyer pays at contract initiation a premium.

Example

An investor wants to buy 1 Mio. EUR call CHF put struck at 1.3000 with the reference rate CHFEUR 1.200. The notional amount in CHF is 1.2 Mio. The option premium can be stated in different forms: Either as a unit or as a percentage and either related to the base or numeraire currency. Therefore, four possible quotations follow. If the premium is quoted in units of the numeraire currency, i.e. the fair price is 0.0075 EUR per one CHF, the premium of the contract is 7'500 EUR. Using percentage of the numeraire the premium is $0.0075/1.200 \times 100 = 0.625\%$ for one unit of CHF. Total premium is $0.625 \times 1'200'000/100 = 7'500$ EUR.

Quotations of FX options are not in term of the premia but in terms of the implied volatility. Strikes are quoted in terms of the option's Delta. Hence the strike level is not fixed before the deal is closed. More precisely, the two parties agree on the FX spot rate and the implied volatility. Then the strike level is fixed at the level such that the Black and Scholes formula yield the Delta value which the two parties are considering. The trades often include the Delta hedge, i.e. a spot trade which offsets the Delta exposure is closed for the transaction. For example a '6m CHF call EUR put 20D 10' means that a trader asks for a CHF call EUR put with 6m expiry, 10 Mio. CHF notional and a Delta CHF of 20 percent. In FX option markets different structures are liquid. A common structure or strategy is the **ATM straddle**: The sum of a call and a put in a base currency. This reflects that an investor's believes in a strong move but is not certain about the direction. ATM has a multiple meaning. **ATM spot** means that the strike is set equal to the FX spot rate. **ATM forward** means that is set equal to the forward rate. A **zero Delta Straddle** has by definition a Delta value zero at maturity: the Delta of the call and put are equal with different sign. An other liquid strategy is the **25D risk reversal**. Such a strategy consists of a long OTM call and a short OTM put in the same base currency. Since traders quote the volatility, the 25 risk reversal is equal to the difference of a 25 Delta call and a 25 Delta put option, i.e. a put where the strike is chosen such that a Delta of -25 percent follows. If the risk reversal is positive the volatility for the call exceeds the put volatility. Therefore the distribution of the expected returns is skewed.

Example

We consider a call spread for FX CHFEUR underlying, i.e. how many CHF are needed to buy EUR. The call spread entitles a company to buy an agreed amount of a EUR at maturity for the given long strike, see Figure 5.11, i.e.

- Company buys CHF call EUR put with lower strike.
- Company sells CHF call EUR put with higher strike.

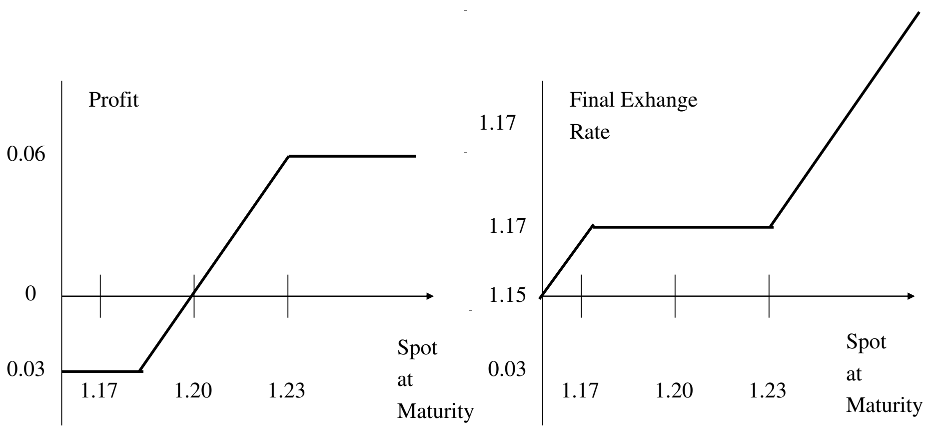


Figure 5.11: Payoff and final exchange rate diagram for a call spread.

If the exchange rate is above the short strike at maturity, the holder's profit is limited to the spread as defined by the short and long strikes. Buying a call spread provides protection against a rising EUR with full participation in a falling EUR. If the underlying value is at maturity below the lower strike value the option is not exercised and the maximum loss, i.e. the loss of the option premium, follows. In this case EUR can be bought at a lower spot in the market. If the underlying value lies between the two strikes, the option is exercised and EUR are bought at the lower strike. If the underlying is below the higher strike, the EUR are bought at a rate $K_2 - K_1 < S_T$, i.e. at a rate below the final underlying rate. To be more specific, consider a Swiss company which wants to hedge receivables from export transaction in USD which are due in 6 months. The company expects the Swiss Franc to become stronger compared to the USD. The company wishes to buy CHF at a lower spot rate compared to the USD if the CHF should become weaker and the company also wishes to be protected against a stronger CHF. A pure call position in CHFUSD is too expensive. Therefore, the company enters in a call-spread position: She buys CHF call USD with a lower strike than the spot reference which we assume to be 1.2000 USDCHF and she sells CHF call USD at a higher strike than the reference spot. She could for example choose 1.1700 for the long call and 1.2300 for the short one.

Consider a CP product for a **German investor**. The zero bond is denominated in Euro and the participation is on NIKKEI 225 Index in Japanese Yen. It is natural to denominate the whole product in Euro. Different approaches are possible to achieve this. Either the price of the underlying asset is translated at the current spot rate both at issue date and maturity date or the change in price is measured in the original currency but then paid out in the bond's issue currency.
Case 1:

$$C_{Euro/Share}(t) = \max\left(E_{Euro/Yen}(t) S_{Yen/Share}(t) - K_{Euro/Share}, 0\right) \tag{5.15}$$

with $E_{Euro/Yen}$ the YENEUR spot exchange rate. Hence, the foreign equity options are **struck in the domestic currency**. The payoff of a call option for the German investor in the domestic currency Euro reads ('F' means 'Foreign', 'D' means 'Domestic'):

$$C_{\mathtt{D}}(t,T) = E_{\mathtt{D\text{-}F}}(t) S_{\mathtt{F}}(t)\, \Phi(d_1) - e^{-r\tau} K_{\mathtt{D}}\, \Phi(d_2) \tag{5.16}$$

where:

$$\begin{aligned}
d_1 &= \frac{\ln(E_{\mathtt{D\text{-}F}} S_{\mathtt{F}}/K_{\mathtt{D}}) + r\tau + \frac{1}{2}\tau\sigma^2_{E_{D-F}S_F}}{\sigma_{E_{\mathtt{D\text{-}F}}S_F}\sqrt{\tau}} \\
d_2 &= d_1 - \sigma_{E_{\mathtt{D\text{-}F}}S_F}\sqrt{\tau}\ ,\ \tau = T - t \\
E_{\mathtt{D\text{-}F}} &= \text{Spot exchange rate in units of Dom per unit of FOR} \\
\sigma_{E_{\mathtt{D\text{-}F}}S_{\mathtt{F}}} &= \sqrt{\sigma^2_{E_{\mathtt{D\text{-}F}}} + \sigma^2_{S_{\mathtt{F}}} - 2\rho_{E_{\mathtt{D\text{-}F}}S_F}\sigma_{E_{\mathtt{D\text{-}F}}}\sigma_{S_{\mathtt{F}}}} \\
&\ \text{Volatility of the underlying } E_{\mathtt{D\text{-}F}} \text{ and } S_F (\text{annualized})\ .
\end{aligned} \tag{5.17}$$

The payoff of a call option for the German investor in the foreign currency JPY reads:

$$C_{\mathtt{D}}(t,T) = S_{\mathtt{F}}(t)\, \Phi(d_1) - E_{\mathtt{F\text{-}D}} e^{-r\tau} K_{\mathtt{D}}\, \Phi(d_2) \tag{5.18}$$

where the changes are the use of $E_{\mathtt{F-D}}$ instead of $E_{\mathtt{D-F}}$ in all parts of the formula. The derivation is based on the change of measure technique developed by Margrabe in 1978.

Case 2:

The next variation is so-called **Quanto options (quantity adjusting option)**, i.e. a fixed exchange rate foreign equity option which are denominated in another currency than the underlying equity exposure. Such products are attractive for investors who wish to have exposure to a foreign asset without the corresponding exchange rate risk. Quantos shield the German investor from EURJPY exchange rate fluctuations. If he invests directly in the NIKKEI, he is exposed to both fluctuations in the NIKKEI index and the EURJPY exchange rate. Essentially, a quanto has an embedded currency forward with a variable notional amount. This variable notional amount gives quantos their name. For the German investor with a call on NIKKEI 225 we have

$$C_{Euro/Share} = E_{P,Euro/Yen} \max\left(S_{Yen/Share} - K_{Yen/Share}, 0\right) \tag{5.19}$$

with $E_{P,Euro/Yen}$ the predetermined exchange rate specified in units of the domestic currency per unit of the foreign currency. The Quanto call option price reads in Euro:

$$C_{\mathtt{D}}(t,T) = E_{P,\mathtt{D-F}}\left(S_F(t)e^{x\tau}\,\Phi(d_1) - e^{-r\tau}K_{\mathtt{F}}\,\Phi(d_2)\right) \tag{5.20}$$

where:

$$\begin{aligned}
d_1 &= \frac{\ln(S_{\mathtt{F}}/K_{\mathtt{F}}) + (r_{\mathtt{F}} - r_{\mathtt{D}})\tau - (\rho\sigma_{S_{\mathtt{F}}}\sigma_{S_{\mathtt{D}}} - \frac{1}{2}\sigma^2_{S_{\mathtt{F}}})\tau}{\sigma_{S_{\mathtt{F}}}\sqrt{\tau}} \\
d_2 &= d_1 - \sigma_{S_{\mathtt{F}}}\sqrt{\tau} \\
\sigma_{S_{\mathtt{F}}} &= \text{Volatility of the underlying } S_{\mathtt{F}} \text{ (annualized)} \\
\rho &= \text{Correlation between asset and DOM exchange rate} \\
x &= r_{\mathtt{F}} - r_{\mathtt{D}} - \rho\sigma_{S_{\mathtt{F}}}\sigma_{\mathtt{D}} \\
E_{P,\mathtt{D-F}} &= \text{ Predetermined exchange rate units of DOM per units of FOR .}
\end{aligned}$$

We compare the quanto call (5.19) with the struck-call (5.15). We assume consider EURCHF. The spot rate $E_{\mathtt{D-F}} = 0.9$ and the predetermined exchange rate $E_{P,\mathtt{D-F}} = 0.8$ are given. The stock S in CHF per share is CHF 60 and the strike is EUR 50. We get in the case of a call which is struck:

$$C = \max(0.9 \times 60 - 50, 0) = 4 .$$

For the quanto call we have.

$$C = 0.8\max(60 - 1/0.8 \times 50, 0) = 0 .$$

Example

Consider a Quanto call with 6 months to expiration. The stock index is 100, strike 105, predetermined exchange rate is 1.5, $r_D = 8\%$, $r_F = 5\%$, volatility of the stock is 20%, volatility of the currency is 10% and the correlation between stock and currency is 0.3. It then follows:

$$d_1 = 0.18711 \ , \ d_2 = 0.04569 \ .$$

Hence,

$$\Phi(d_1) = 0.42578 \ , \ \Phi(d_2) = 0.48177 \ ,$$

and the price of the call is

$$\begin{aligned} C(0) &= 1.5\left(100e^{0.05-0.08-0.3\times0.2\times0.1}\Phi(d_1) - 105e^{-0.08\times0.5}\Phi(d_2)\right) \\ &= 1.5\,(100\times0.98216\times\Phi(d_1) - 105\times0.67032\times\Phi(d_2)) \\ &= 11.8646 \ . \end{aligned}$$

It is worth to understand quanto modelling. Consider a gold contract in XAU-USD quotation. The goal is to quanto this contract into EUR (=DOM) using Black and Scholes. EUR is the numeraire currency. Gold and USD rates versus EUR are modelled as:

$$\begin{aligned} \texttt{XAU-USD}: \quad S^1 &= S^3/S^2 \quad \left(= \frac{\text{XAU}}{\text{EUR}}\frac{\text{EUR}}{\text{USD}}\right) \\ \texttt{XAU-EUR}: \quad dS^3/S_3 &= (r_{\text{EUR}} - r_{\text{XAU}})dt + \sigma_3 dW^3 \\ \texttt{USD-EUR}: \quad dS^2/S_2 &= (r_{\text{EUR}} - r_{\text{USD}})dt + \sigma_2 dW^2 \end{aligned}$$

$dW^3dW^2 = -\rho_{23}dt$ where a minus sign in front of the correlation follows because both processes have the same base currency, see Figure 5.12. The goal is to find the dynamics of dS^1/S^1. The dynamics follows from Itô's formula. It follows that for triangular FX-markets, i.e. the two currencies XAU and USD have EUR as common currency in our example, correlations between currencies can be expressed in terms of volatilities. Consequently we do not need to estimate correlation coefficients and we can hedge correlation risk merely by trading volatility. If we calculate the variance of the logarithm for $S^1S^2 = S^3$ we get

$$\rho_{12} = \frac{\sigma_3^2 - \sigma_1^2 - \sigma_2^2}{2\sigma_1\sigma_2} \ .$$

But this is the content of the cosine theorem:

$$c^2 - a^2 - b^2 = -2ab\cos\gamma \ .$$

This calculation can be visualized using elementary geometry, see Figure 5.12. The label of the triangle are the currencies. If volatilities are represented as edge vectors with the length of the vectors equal to the value of the volatility, the law of cosine states that the cosines of angles are the correlations.

The elementary geometry also applies to currency pairs which do not have a common currency. We use this interpretation in our example. Applying Itô's formula one gets[2]

$$dS^1/S^1 = (r_{\text{USD}} - r_{\text{XAU}} + \sigma_2^2 + \rho_{23}\sigma_2\sigma_3)dt + \sigma_3 dW^3 + \sigma_2 dW^2 \ .$$

[2] $dS^1/S^1 = d(S^3/S^2)S^2/S^3$. But

$$d(S^3/S^2) = dS^3/S^2) - S^3dS^2/(S^2)^2 + dS^3/d(S^2).$$

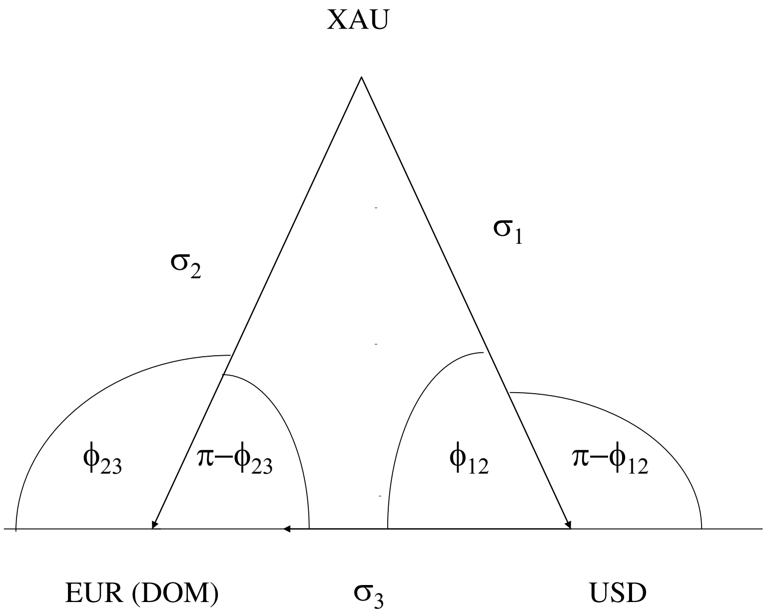


Figure 5.12: XAU-USD-EUR Quanto Triangle. The arrows point in the direction of the respective base currencies. The length of the edges represents the volatility. The cosine of the angles $\cos\phi_{ij} = \rho_{ij}$ represents the correlation of the currency pairs i and j, if the base currency (DOM) of S^i is the underlying currency of S^j (FOR). If both S^i and S^j have the same base currency (DOM), then the correlation is denoted by $\cos(\pi - \phi_{ij}) = -\rho_{ij}$.

Since S^1 is a geometric Brownian motion with volatility σ_1, we introduce a new Brownian motion W^1, i.e.

$$dS^1/S^1 = (r_{\text{USD}} - r_{\text{XAU}} + \sigma_2^2 + \rho_{23}\sigma_2\sigma_3)dt + \sigma_1 dW^1 \ .$$

The geometry in the Quanto Triangle in Figure 5.12 and the cosine law imply

$$\sigma_3^2 = \sigma_1^2 - 2\rho_{12}\sigma_1\sigma_2 \ , \ \sigma_1^2 = \sigma_2^2 + 2\rho_{23}\sigma_3\sigma_2$$

with ρ_{12} the correlation between XAU-USD and USD-EUR. These equations imply

$$\sigma_2^2 + \rho_{23}\sigma_2\sigma_3 = \rho_{12}\sigma_1\sigma_2 \ .$$

Inserting this in the dS^1/S^1-dynamics gives

$$dS^1/S^1 = (r_{\text{USD}} - r_{\text{XAU}} - \rho_{12}\sigma_1\sigma_2)dt + \sigma_1 dW^1$$

which is the drift adjusted process leading to the pricing formula (5.16). More precisely, we have the correspondence:

- $r_{\text{F}} - r_{\text{D}} \rightsquigarrow r_{\text{USD}} - r_{\text{XAU}}$.
- $\rho\sigma_{S_F}\sigma_{S_D} \rightsquigarrow \rho_{12}\sigma_1\sigma_2$.
- $\sigma_{S_F} \rightsquigarrow \sigma_1$.

Case 3:

A different motivation is obtained by using **composite options** instead of quanto options. Whereas in quanto options the investor gets a return regardless of the exchange rate the buyer of composite option has an exposure to exchange rate. The reason to use composite options is for example to protect a value in the investor's own currency on a foreign investment. To specify this consider a US investor which owes Nestle stocks denominated in CHF.

- The price of the stocks today is CHF 50 and the exchange rate is 0.8 USD for 1 CHF. Hence the dollar value of one Nestle share is USD 40.
- After one year Nestle went down in Swiss currency to CHF 45 and the exchange rate decreased to 0.5 USD for 1 CHF, i.e. the dollar became stronger relative to the Swiss Franc. The dollar value of a Nestle share has gone down from USD 40 to USD 22.5.

The last expression is equal to:

$$dS^3/d(S^2) = -\frac{1}{(S^2)^2}dS^3dS^2 = -\frac{1}{(S^2)}\rho_{23}\sigma_2\sigma_2 dt.$$

Inserting the dynamics proves the formula.

- To protect the dollar value the investor buys ATM composite put options, i.e. the strike price is USD 40. The share price loss is offset by the composite put option which is worth the difference USD 40-22.5. The holder of the option therefore protected his Nestle investment both from exchange rate and stock price movements.

Example:
To summarize, we compare three types of cross currency options: Quanto, composite and foreign currency options. The underlying values are gold (XAU) measured in USD and CHF. The option payoff reads:

1. Quanto: $N \max\left(\frac{\mathrm{XAU}_{\mathrm{USD}}(T)}{\mathrm{XAU}_{\mathrm{USD}}(0)} - \%K, 0\right)$.
2. Composite: $N \max\left(\frac{\mathrm{XAU}_{\mathrm{USD}}(T)/\mathrm{CHFUSD(T)}}{\mathrm{XAU}_{\mathrm{USD}}(0)/\mathrm{CHFUSD}(0)} - \%K, 0\right)$.
3. Foreign Currency: $N \frac{\mathrm{CHF}}{\mathrm{USD}}(T) \max\left(\frac{\mathrm{XAU}_{\mathrm{USD}}(T)}{\mathrm{XAU}_{\mathrm{USD}}(0)} - \%K, 0\right)$.

We assume

- at time 0: XAU = 1500 USD, CHF=0.9 USD, a notional N of 1 Mio. CHF and an ATM option.
- at time T: Consider rising gold prices and the Swiss Franc gets stronger compared to the USD, i.e. XAU(T) = 1700 USD and CHF=1 USD at time T.

Consider rising gold prices and the Swiss Franc gets stronger compared to the USD, i.e. XAU(T) = 1700 USD and CHF=1 USD at time T. We get for the three options at time T:

$$\text{Quanto :} \quad 1 \times \max\left(\frac{1700}{1500} - 1, 0\right) = 133'333 \text{ CHF}$$

$$\text{Composite :} \quad 1 \times \max\left(\frac{1700/1}{1500/0.9} - 1, 0\right) = 0 \ \text{ CHF}$$

and

$$\text{Foreign Currency :} \quad 1 \times 0.9/1 \times \max\left(\frac{1700}{1500} - 1, 0\right) = 120'000 \text{ CHF} .$$

If gold falls and the Swiss franc gets weaker against the USD, i.e. XAU(T) = 1300 USD and CHF=0.75 USD at time T. We get for the three options at time T:

$$1 \times \max\left(\frac{1300}{1500} - 1, 0\right) = 0 \ , \ 1 \times \max\left(\frac{1300/0.75}{1500/0.9} - 1, 0\right) = 40'000 \ ,$$

and

$$1 \times 0.9/0.75 \times \max\left(\frac{1300}{1500} - 1, 0\right) = 0 \text{ CHF} .$$

The foreign currency option are not very common (why?). In this case Quanto style is meaningful for an investor with a view on XAU without carrying about FX movements.

The composite case is suitable for a CHF investor who already has an XAU exposure and who wants to protect the CHF value of his portfolio.

We described several possibilities to increase the participation rate. The results so far were only qualitative. Table 5.4 gives some quantitative figures.

Change in	Impact on Part. Rate	Approx. increase in Participation
Dividend yield	Almost linear	10% per 1% change in div. yield
Basket correlation	Almost linear	2% per 0.1 change in pos. correlation
Foreign reference index	Almost linear	7% per 1% change in $r_{\mathbf{F}}$
Cape Rate	Piecewise linear	2% per 10% change in cap rate from 90% – 140%
"	Piecewise linear	10% per 10% change in cap rate from 70% – 90%
"	Piecewise linear	120% per 10% change in cap rate from 50% – 60%

Table 5.4: Approximative quantitative impact of variations in the CP design on the participation rate. Values for the volatility used is 20% and interest rate p.a. of 3 percent.

5.2.5.1 Returns and Indices

Investor typically think in terms of return on investment. The notion of quanto and composite options allow us to defines several return types. The simplest return for an investment starting at time 0 up to time T is given by the simple return

$$R_{0,T} = \frac{I_T - I_0}{I_0} \ . \tag{5.21}$$

Suppose that the index is denominated in a different currency than the investor is used to and we have to do a currency conversion. The first method, called quanto, uses the known exchange rate E_0 at time 0 to convert all cash flows we get the return on a quantoed index

$$R^q_{0,T} = \frac{E_0 I_T - E_0 I_0}{E_0 I_0} = \frac{I_T - I_0}{I_0} \ . \tag{5.22}$$

If we instead use the spot exchange rates we get the return on a composite index:

$$R^c_{0,T} = \frac{E_T I_T - E_0 I_0}{E_0 I_0} \ . \tag{5.23}$$

Consider the DAX with values $I_0 = 6'000$ and $I_T = 7'000$. The currency pair is EURCHF with $E_0 = 1.2$ and $E_T = 1.3$. Then,

$$R_{0,T} = R^q_{0,T} = 1/6 = 16.6\% \ , \ R^c_{0,T} = \frac{1.3 \times 7'000 - 1.2 \times 6'000}{1.2 \times 6'000} = 26.4\% \ .$$

So far we considered a single index. If we use a domestic basket, i.e.

$$B_t = \sum_{j=1}^{n} w_j I_j(t)$$

of n indices the simple return on domestic basket $BR_{0,T}$ is equal to the weighted sum of simple index returns:

$$BR_{0,T} = \sum_{j=1}^{n} x_j R_{j,0,T}. \tag{5.24}$$

with $x_j = w_j I_j(0)/B(0)$. The same holds true for a basket of quantoed indices. If one uses the basket as a reference index, i.e. the basket value $B_r(0) = 100$ is set equal to 100 at time zero, the basket value at time t reads

$$B_r(t) = B_r(0) \sum_{j=1}^{n} x_j \frac{I_j(t)}{I_j(0)} .$$

The simple return on the basket is given by

$$BR_{0,T} = \frac{B_r(T) - B_r(0)}{B_r(0)} . \tag{5.25}$$

The same applies to the quanto or composite basket case, see Figure 5.13.

So far the only dates which matter for the return where initial date and a future date T. We next want a return which is based also on intermediate values of the indices. Consider a single index. Suppose that there are m index values which matter for the return calculation. The average index value is

$$\overline{A_T} = \frac{1}{m} \sum_{k=1}^{m} I(m)$$

and the average return on a domestic index is defined by:

$$\overline{R_{0,T}} = \frac{\overline{A_T} - I_0}{I_0}. \tag{5.26}$$

Algebra gives us

$$\overline{R_{0,T}} = \frac{1}{m} \sum_{k=1}^{m} R_{0,k} , \tag{5.27}$$

i.e. the average return on a domestic index is equal to the average of simple returns. The same holds true for the two currency conversion types. We note that the basis is always I_0 which is compared to $I(1), I(2), \ldots, I_T$. If we generalize to baskets of indices the above results carry over. Figure 5.13 shows the impact of averaging. The index is assumed to increase in the first time steps and then to drop in the last step back to its initial value. The average index value and the average return are then calculated. The average return on the index is 22.9 percent compared to the zero return for the whole period if we use simple return calculation. This shows the stabilizing effect of averaging - risk for an investor that a drop at the end of the contract destroys the former performance is reduced at the price that a bust at the end of the period also has a small impact.

	Return on Domestic Basket			
	Index 1	Index 2	Index 3	
Weight w	0.5	0.25	0.25	
Time 0	100	100	100	
Time 1	120	130	90	
Time 2	115	140	70	
	Return Index 1	Return Index 2	Return Index 3	Return Basket
Time 1	20%	30%	-10%	0.15
Time 2	-4%	8%	-22%	-0.06

Average Index Values	
	Index Value
Time 0	100
Time 1	110
Time 2	120
Time 3	130
Time 4	145
Time 5	155
Time 6	100
Average Index Value	122.9
Average Return Index Value	22.9%
Total Period Return	0%

Figure 5.13: Examples for returns on a domestic basket and average index values. The basket is not equally weighted but the first index has double weight compared to the two other ones.

Proposition 5.2.2. *The average return on a basket of domestic indices*

$$\overline{BR_{0,T}} = \frac{\overline{B(T)} - B(0)}{B(0)} \tag{5.28}$$

with

$$\overline{B(T)} = \frac{1}{m}\sum_{k=1}^{m} B_k(T)$$

is equal to the weighted sum of the average returns on the component domestic indices

$$\sum_{k=1}^{m} x_k \overline{R_{k,0,T}}$$

with $\overline{R_{k,0,T}}$ *the average return on the domestic index* k.

The content of the proposition also holds for the currency conversion cases.

Instead of averages one can use extreme values to calculate returns. Let

$$I^* = \max(I(1), \dots, I(m))$$

be the maximum of an index at fixed dates $1, 2, \dots m$. The maximum simple return $MR_{0,T}$ is defined as:

$$MR_{0,T} := \frac{I^* - I_0}{I_0} = \max(R_{0,1}, R_{0,2}, \dots, R_{0,m}) \ . \tag{5.29}$$

An extension of the extreme value returns is as follows: Compare the index value with prefixed set of reference values $H^1, \dots, H^l$ (ladders) at fixed monitoring points instead with the index value itself. To obtain the cash flows at a maturity date T and a return formula we use the indicator function:

$$\chi_j = \begin{cases} 1, & \text{if there exists a monitoring date } t \in \mathbb{T} : I(t) \geq H_j; \\ 0, & \text{else.} \end{cases} \tag{5.30}$$

with $\mathbb{T}$ the set of monitoring dates. A cash flow at time T is given by

$$CF(T) = \max(I_0, \chi_1 H^1, \dots, \chi_l H^l) \ .$$

The stepwise maximum return $\mathbb{SR}_{0,T}$ then reads:

$$\begin{aligned} SR_{0,T} &= \frac{\max(I_0, \chi_1 H^1, \dots, \chi_l H^l) - I_0}{I_0} \\ &= I_0 \frac{\max(1, \frac{\chi_1 H^1}{I_0}, \dots, \frac{\chi_l H^l}{I_0}) - I_0}{I_0} \\ &= \max(1, \frac{\chi_1 H^1}{I_0}, \dots, \frac{\chi_l H^l}{I_0}) - 1 \\ &= \max(0, \frac{\chi_1 H^1}{I_0} - 1, \dots, \frac{\chi_l H^l}{I_0} - 1) \\ &= \max(0, \chi_1 \frac{H^1 - I_0}{I_0}, \dots, \chi_l \frac{H^l - I_0}{I_0}) \ . \end{aligned} \tag{5.31}$$

5.2.6 Approximation Formulae

We consider some approximation formulae in derivative and RSP pricing. The approximations allow one for quick estimation of prices and Greeks of options. The formulae are based on the approximation of the standard normal distribution, i.e. we assume Black and Scholes as pricing tool.

We start with zero interest rates and zero dividend yield: **r=d=0**. Then ATM call and put prices are the same and approximated by

$$C_{\text{app}}(0,S,S,T) = P_{\text{app}}(0,S,S,T) = 0.4\sigma S\sqrt{T}$$

where $0.4 \sim \frac{1}{\sqrt{2\pi}}$. The equivalence of put and call prices follows from the Put-Call parity:

$$P = e^{-rT}K - S + C = S - S + C \ .$$

The approximation follows from Black and Scholes formula. We have

$$C = S\Phi(\frac{\sigma\sqrt{T}}{2}) - S\Phi(-\frac{\sigma\sqrt{T}}{2}) \ .$$

Since $\Phi(-\frac{\sigma\sqrt{T}}{2}) = 1 - \Phi(\frac{\sigma\sqrt{T}}{2})$ we get

$$C = 2S\left(\Phi(\frac{\sigma\sqrt{T}}{2}) - 1/2\right) \ .$$

A Taylor series expansion for the normal distribution reads

$$\Phi(x) = \Phi(0) + \Phi'(0)x + O(x^2) = 1/2 + \frac{x}{\sqrt{2\pi}} + O(x^2) \ .$$

Thus

$$C = 2S\left(\Phi(\frac{\sigma\sqrt{T}}{2}) - 1/2\right) \sim 0.4\sigma S\sqrt{T} \ .$$

This approximation allows us also to solve for approximative implied volatility instead.

The approximation of an European call option ATM in the Black and Scholes model with $r, d \neq 0$ is derived in the same way and reads:

$$C_{\text{app}}(0,S,S,T) = \sigma S\sqrt{\frac{T}{2\pi}}(1 - \frac{(r+d)T}{2}) + \frac{(r-d)T}{2}S \ . \tag{5.32}$$

In the same way approximate Greeks can be derived. The Delta of a ATM call with $r = d = 0$ is

$$\Delta_{\text{app}}(C) = 1/2 + 0.4d_1 = 1/2 + 0.2\sigma\sqrt{T}$$

and the Vega is

$$\text{Vega}_{\text{app}} = C_{\text{app}}(0,S,S,T)/\sigma \ .$$

What is the sensitivity of a call if interest rates are not zero and if they increase by one percent? If rates increases borrowing will become more expansive and therefore the price for the amount needed to buy the Delta for the call increases. But option prices are also discounted prices. Therefore an increase of the interest rates reduces the option prices due to discounting. The extra costs from borrowing are

$$\Delta_{\mathtt{app}} \times \text{Change Interest Rate}$$

and the reduction in discounting is given by

$$\text{Change Interest Rate} \times \text{Time to Maturity} \times \text{Option Price}_{\mathtt{app}}$$

Example
Consider a ATM call with maturity one year and implied volatility $\sigma = 32\%$. Then approximative Vega is equal to $0.4\sqrt{1} = 0.4$, i.e. 40 bp, for a one percent movement in volatility. The approximative call price is $0.4 \times 32\% = 12.8\%$. The approximative Delta then follows: $\Delta_{app} = 0.564$. The approximative impact of an one percent interest increase is

$$\Delta_{\mathtt{app}} \times \text{Change Interest Rate} = 56.4 \times 1\% = 56.4 \ bp$$

minus the discounting effect

$$\text{Change Interest Rate} \times \text{Time to Maturity} \times \text{Option Price}_{\mathtt{app}} = 1\% \times 1 \times 12.8\% = 12.8bp$$

i.e. 43.6 bp is the impact of a one percentage increase of the interest rate. This is the approximative Rho sensitivity. The new price of the call, which is by Taylor Series

$$C_{\mathtt{ITM/OTM,\ app}} = C_{\mathtt{ATM,\ app}} + \text{Sensitivity} \ \times \delta\text{Parameter},$$

is given by

$$C = 12.8\% + 0.436 \times 1 = 13.236 \ .$$

Note that in this approximation of an OTM call using the ATM call as zeroth order starting point no volatility smile and volatility skew is used.

We use these approximations to derive approximative formulae for capital protected products. In general we have for a CP the decomposition

$$V_T = af_{\mathtt{prot}}(a, b, \cdots) + bf_{\mathtt{part}}(a, b, \cdots)$$

into a capital protection and participation part. The participation part, i.e. sloppy the options, are non-linear functions of the participation rate b and other parameters. Therefore, it is not possible in general to solve the above equation explicitly w.r.t. to participation b or other variables of interest rate such as a coupon value or a capital protection level a. Using the above approximation of the normal distribution we are able to say obtain explicitly a participation rate as a function of all other parameters.

Using the replication formula for a CP product without coupon and without cap we have for the approximative participation rate:

$$b_{\text{app}} = S \frac{100\% - \frac{\text{KL}}{(1+r)^T}}{C_{\text{app}}(0, S, S, T)}$$

with KL the capital protection level. The last equation can also be used to write the approximate capital protection level as a function of the participation rate. Table 5.14 shows the approximative participation rate.

	T=1	**r=2%**							
KL	**100**	**97.5**	**95**	**92.5**	**90**	**87.5**	**85**	**82.5**	**80**
Vol 20%	22%	50%	77%	105%	132%	160%	187%	215%	242%
Vol 30%	15%	34%	53%	72%	92%	111%	130%	149%	168%
	T=3	**r=2%**							
KL	**100**	**97.5**	**95**	**92.5**	**90**	**87.5**	**85**	**82.5**	**80**
Vol 20%	35%	50%	64%	78%	93%	107%	121%	136%	150%
Vol 30%	25%	35%	45%	56%	66%	76%	86%	96%	107%
	T=3	**r=4%**							
KL	**100**	**97.5**	**95**	**92.5**	**90**	**87.5**	**85**	**82.5**	**80**
Vol 20%	58%	70%	82%	94%	105%	117%	129%	140%	152%
Vol 30%	44%	52%	61%	70%	78%	87%	96%	105%	113%

Figure 5.14: Relationship between different parameters for a CP product without cap and without coupons. The table shows the approximative participation rate. Dividend yield is $d = 0$.

We apply the approximation method to the case where the CP product has a cap. In this case an option OTM is needed. If volatility is constant, this would be not problematic. But in fact, there is a volatility smile and/or a skew. A simple method to account for this is as follows. Given an underlying value - here the SMI - we use a regression in the variables moneyness and maturity. That is, we estimate the ratio

$$\text{Ratio} = \frac{\text{Price OTM}}{\text{Price ATM}}$$

for a volatility constant at 20%. The regression implies

$$\text{Ratio} = 2.93 - 2.22\text{Moneyness} + 0.067\text{Maturity} .$$

The estimate has an $R^2 = 0.953$. Using this estimate we get

$$b_{\text{app}} = S \frac{100\% - \frac{\text{KL}}{(1+r)^T}}{C_{\text{app}}(0, S, S, T)(1 - 2.93 - 2.22\text{Cap} + 0.067T)}$$

with Cap the cap level. Contrary to the case with no cap this approximation is market dependent. That is for a given underlying the validity of the ratio has to be tested before one uses the approximation.

Example

We test the quality of the above approximations. For a CP product without cap on SMI we have the parameters: $\sigma = 18\%, r = 1.61\%, d = 2.00\%, T = 5, KL = 95\%$. Using the approximation formula we get

$$b_{\text{app}} = 90.16\%$$

compared to the Bloomberg participation rate of 90.60%. For a CP product with cap on SMI we have the parameters: $\sigma = 21\%, r = 1.14\%, d = 2.5\%, T = 3, Cap = 115\%, b = 110\%$. Using the approximation formula we get

$$\text{KL}_{\text{app}} = 98.36\%$$

compared to the Bloomberg value 98.84%, see Figure 5.15 for a screen shot.

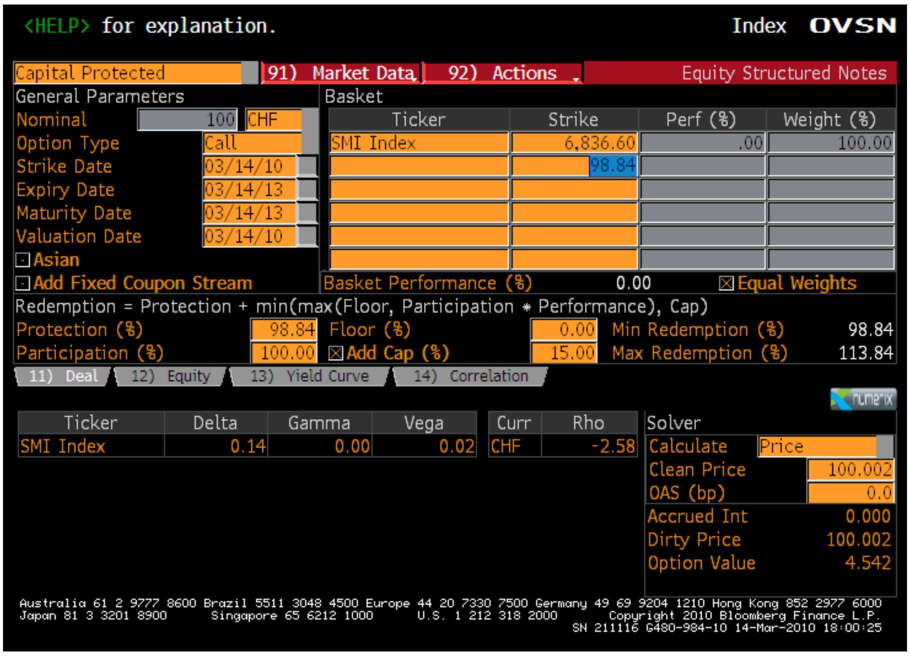


Figure 5.15: Bloomberg solution for a CP product with cap on SMI we have the parameters: $\sigma = 21\%, r = 1.14\%, d = 2.5\%, T = 3, Cap = 115\%, b = 110\%$.

5.3 Trading, Market Making

Since RSP are contracts about futures payoff profiles, traders play a key role in the fulfillment of the contracts. They determine the price of the product and offer a secondary market. The price of a SRP can be posted in two ways. Either the price is given per share, like for stocks, or the price is given in a percentage value of the notional amount. The first method is used for discount-, bonus- and tracker products and also for warrants. The second one, which corresponds to the pricing of bonds, if used for barrier reverse convertibles and capital protected products.

After issuance, the issuer faces market risk in the option parts of the products. There are two ways to handle them.

Buy the components
The issuer buys the components of the product from derivative houses, i.e. firms which are specialized to deal with risks of derivatives which the issuer does not want to manage or is not able to do so. Investment banks are such specialized derivative firms. The contract between the issuer and the derivative firm is an OTC-contract. The advantage of this method, also called white-labelling, for the issuer is the elimination of market risk. But the issuer faces credit risk of the derivative house and a part of the value of chain is lost to the specialized firm.

Management in the own trading book
This approach allows to keep the whole value chain in-house. The disadvantage are the need for a proper risk management process together with the costs for such a process (know how and IT are the key cost drivers). In practice one observes two different organization forms of the trading units. Either they are organized according to the product categories or they are organized following the underlying values. Trading desk which are specialized on their products cover many underlying values. The advantage is that pricing is coherent for a comprehensive trading unit and people have a clear focus on the products. A disadvantage is an only superficial knowledge about underlying values and basis risk of the products, i.e. the risk that the same product has different prices and different risk sensitivities on different markets.

How does a trader hedges the risks? Often one encounters two type of traders - flow traders and back-book traders. Flow traders act as market maker in the secondary market. They hedge the simplest risks such as the Delta exposure. More complicated risk factors are considered in the back books. Traders are measured how they manage their risk and also their earning objectives. How does traders generate profits? First, if the fair value of SRP is say 98 percent of the notional amount an issuance margin of 2 percent is earned if the product is issued at par. Theoretically, if a perfect hedge exists traders could lock in this 2 percent. Often such perfect hedges are not feasible and the initial margin is at risk. Furthermore, traders are not forced to perfect hedge - they can play their view within the top-down allocated risk limits. Today, trading based on client flow is dominating proprietary trading where trading is not based on any client transactions, i.e. trading is based on the bank's own equity.

We consider how volatility trading on equities can generate profit and loss (P& L). We have to distinguish between vanilla derivatives such as warrants and exotic options such as DIP in barrier reverse convertibles. The goal is to hedge the DIP using also vanilla options. There are different methods to do this - static and dynamic ones. We assume that a DIP can be written as linear combination of vanilla options. Hence, the trader of structured products buys vanilla options from the vanilla option trader for clients buying barrier reverse convertible. The structured product trader therefore sell DIPs. The warrants trader sells the same type of options to vanilla option investors. At this point one observes that the structured product trader buys the option at a lower pricer then the warrant trader is selling the same option, where the difference is larger than the bid-ask spread. How is this possible? To understand this, we recall that **implied volatility** σ_{im} is the key parameter value which put into the theoretical pricing formula equalizes the market price:

$$\text{TheoPrice}(\sigma_{\text{im}}) = \text{MarketPrice} \; .$$

This requires an option pricing model such as the Black and Scholes model for example. We know that implied volatility is not constant: There is a smile and a skew.

The traders use the calculated volatility curves from say Eurex options prices. Given the mid-curves between bid and ask, the structured product trader will value the volatility below the mid-curve and the warrants trader above this curve. This defines a risk less profit for both trading desk on a stand alone basis and integrated for the whole trading floor. Arbitrage, i.e. a risk less strategy to exploit this fact, is not possible since it is not possible to be short the SRP. The market is in this sense incomplete. Therefore the SRP trader can sell the SRP with a lower volatility than the Eurex-induced value. Is there a guarantee that the SRP is not priced way-off its fair value? Competition between the issuers, which is strong in particular in Switzerland, and the existence of online tools to price the SRP force the issuer to a competitive price behavior.

If we consider Black and Scholes pricing for barrier option a problem with this model shows up. Consider a vanilla option DAX and a Down& In Put (DIP) barrier option, see Figure 5.16.

In panel A suitability of the Black and Scholes model for the vanilla option is shown. For a Put with ATM strike we would use a volatility of 25 percent. The price would be in-line with the market. Panel B shows the difficulty if we consider the same model for the DIP: What volatility to use for a Down& In Put with ATM strike and barrier say 50 percent of the strike? A huge volatility range follows. If one considers the price of the DIP using the two extreme volatilities, one gets prices of 23 respectively 369. Extensions of the Black and Scholes models such as the Local Volatility (LV) or stochastic volatility models address these issues.

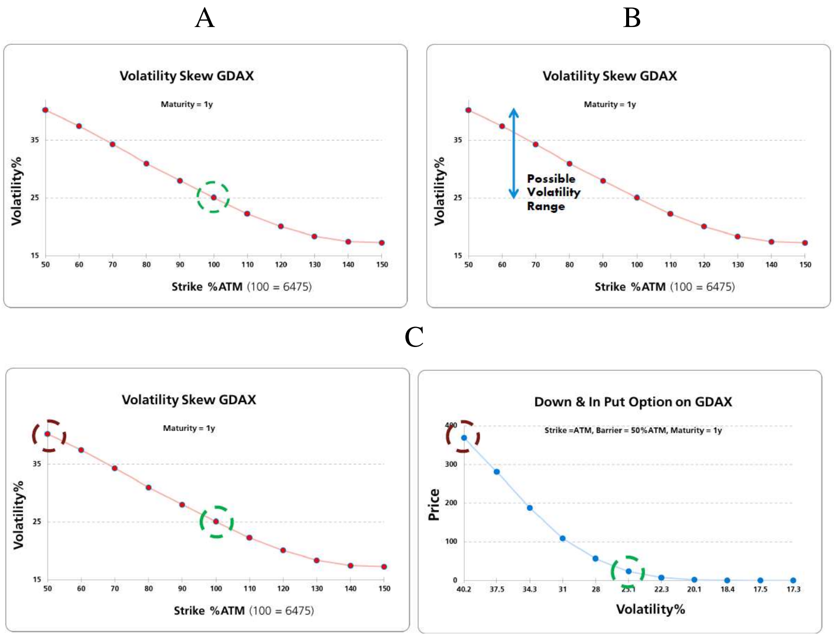


Figure 5.16: Local volatility: The problem of Black and Scholes. *Source: K. Navaian, Zurich Cantonal Bank, 2012.*

Chapter 6

Less Traditional Asset Classes / Investments

6.1 Real Estate (RE)

6.1.1 Economics, Products

Real estate is in terms of value one of the largest asset classes. In most countries the value of the houses is a multiple of the value of all listed stocks. But one also observes very poor financial markets where real estate risk can be restructured and transferred. We show how financial innovation can overcome the problem that each house is unique in the sense that no meaningful index for the market price of real estate can be constructed.

The value of all Swiss building is about five times larger than the value of all listed Swiss stocks. Contrary to the size proportion liquidity behaves: The turnover of the broad Swiss stock market index SPI is on average around 100% per annum of its value. The turnover on the residential housing market in Switzerland is around $1-2$ percent per year. Similar figures hold for many other countries too.

The real estate asset class has many characteristics which make it a difficult to define financial markets similar to equity or FX markets. Real-estate (derivatives) markets are still in most countries in a state of infancy. One also observes a great deal of variation across countries (or even within countries) in prices, market trends, institutional settings, market practice regulation and taxation. Why is property illiquid? First, real estate is a durable good, i.e. no frequent trading is needed. Second, property is heterogeneous, i.e. it is difficult to trade since a normalization is needed which makes the different assets comparable. Third, property is indivisible, i.e. the size of the trades are large. This makes trading costly because of the search costs (heterogeneity), transaction costs (legal costs, taxation) and financing costs. Transaction costs of privately held housing are between 8 and 10 percent in Switzerland. Real estate is heavily taxed in most jurisdictions.

Reasons are the simplicity to monitor the real estate assets and the inelastic demand for housing. In many countries, taxation is skewed in favor of owner-occupied housing to the disadvantage of renting (CH as an exception). But real estate markets are also heavily regulated, i.e. rent controls, land use planning, zoning, preservation of historical buildings, etc.

How are house prices determined? First, prices are not readily available although this is changing. Second, prices depend on size, quality and locational considerations which is a rational for experts in the housing markets: valuators, buying agents, selling agents, structural engineers etc. reduce the uncertainty costs of real estate transactions. In the US and the Netherlands the agents act on both sides of the markets, in Australia auctions mechanism apply and in Switzerland the major rule is direct bargaining. Many practitioners follow the wrong belief that segmentation in the real estate market is complete, i.e. each object defines a market due to the differences in quality and locations. But there exists a well-defined relationship between the house price and it's attributes even though theses attributes are bundled in the house and cannot be traded separately. That is one can evaluate the expectation of the price of a house conditional on its attributes. This is called the **hedonic** valuation method.

A first classification in the RE asset class is to differentiate three basic types of property.

- **Residential property**. This consists of condominium apartments and single family houses.
- **Rental property.**
- **Commercial property.**

The three types possess different risks, cash flows and return properties. We consider **only residential property** in the sequel.

A different classification of RE in the dimensions 'markets' and 'asset type' is shown in Figure 6.1.

Investment in real estate can be further classified into **direct**, **indirect** and **derivative** investments. While direct investments choose a specific building, indirect investments invests are of a fund or a Real Estate Investment Trust (REIT) type. REITs (Real Estate Investment Trusts) offer publicly traded common stock shares in companies that own properties and mortgages. REITs are exempted from corporate income tax under certain conditions. They can be closed, open or semi-closed as in CH, where they are quoted on the stock exchange. REIT's are the counter part to mutual stock funds. Figure 6.2 gives an overview over the risk, return and liquidity characteristics of direct and indirect investments. Derivatives are based on real estate indices, see Figure 6.3: The risk and return profile of such an index is structured.

	Public markets	**Private markets**
Equity assets	• RE funds • RE stocks • REITs	• Real property • House equity • Private RE equity
Debt assets	• MBS • Bonds of RE companies • Money instruments	• Bank loans • Whole mortgages • Venture debt

Figure 6.1: Real estate assets. *Source: Geltner and Miller, 2001*

	Direct	**Indirect**		
	In-house portfolios	Non-listed Closed-end funds	Non-listed Open-end funds	REITs
Risk	Middle	High	Low	Middle
Return	Low	High	Low	Middle
Liquidity	Low	Low	Middle	High

Figure 6.2: Direct and indirect investments.

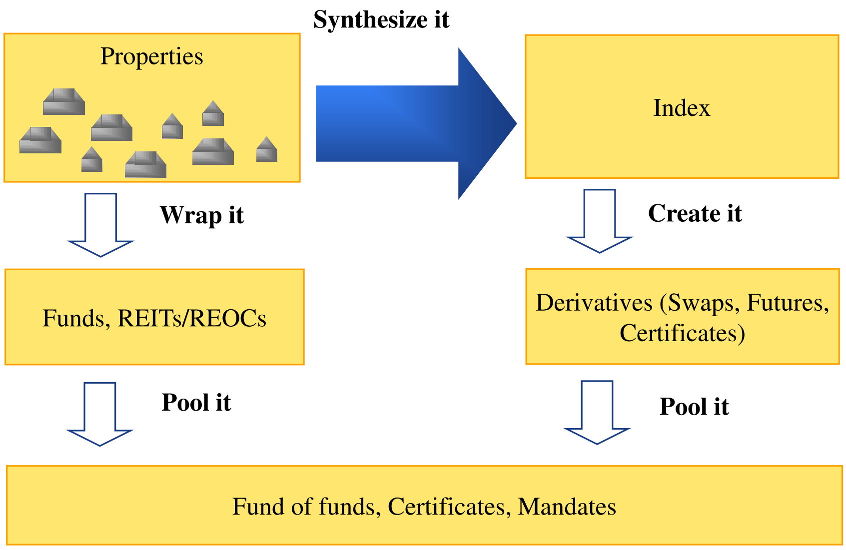


Figure 6.3: Scheme for the fund and derivative structuring of real estate. *Source: Helmuth Aberer, UBS, Swiss Finance Institute Annual Meeting, 2006.*

The three different types of investment described above - direct, indirect, derivative - have different characteristics. Transaction, tax and administration costs are high for direct investments. They are lower in the indirect approach and the lowest one for derivatives. Time to market is long if one considers direct investments and considerably shorter for the two other types. Hedging and short selling is not possible using the direct or indirect approach. Derivatives at least partially offer this - the extend depends on how mature and liquid the derivative markets are. Cherry picking is best achieved following a direct investment, less in indirect ones and almost not possible by applying derivatives. The premium over net asset value is advantageous for direct investments, less for derivatives and a disadvantage if one buys funds or REITs. Model risk is highest at present in the non liquid derivative markets.

Risks of RE investments consists of

- Market risk, i.e. systematic real estate risk which cannot be diversified away within the RE asset class. This risk has several components among them are price -, rent - and vacancy risk. This risk can be measured by housing price indices.

- Idiosyncratic or specific risk. This risk is diversifiable within portfolios.). This risk is difficult to measure. Evidence from repeated-sales indicates that this risk is twice as big as systematic risk.

- Liquidity risk which is at least partially systematic. Low liquidity means also that information is scarce. This risk is also difficult to measure. The risk becomes substantial during downturns.

From a risk perspective the three approaches behave different. Direct investment means to search for the alpha. The focus is on idiosyncratic risk and not on systematic real estate market risk. Such investments are often of a buy-and-hold type due to the missing liquidity, lack of transparency and the high transaction costs. Derivatives on the other extreme focus on systematic real estate risk - that is risk which is directly and exclusively related to the real estate and tenant's market. The underlying value reflects systematic price risk, vacancy risk or rental price risk. The rationale that derivatives do not focus on idiosyncratic risk are moral hazard, monitoring costs and that in larger portfolios idiosyncratic risk diversifies to a certain extend. Indirect investments are somewhere in between the two extremes. They face neither pure idiosyncratic risk, since they are based on a portfolio of buildings, neither pure market risk since business and operational risk of the real estate firms in the portfolio interfere.

6.1.2 Hedonic Indices

We consider real estate indices. We observe two type of indices, **appraisal based** and **transaction based** indices. The first one, such as the IPD indices, are based on the appraisal of a large sample of price appraisers. The second one, such as the Halifax indices for UK together with its sub-indices or the ZWEX for the greater Zurich Area, are based on real transactions. We consider the ZWEX.

This index is generated by thousands of transactions each year. The houses as underlying assets are of **different quality**. Quality means the number of rooms, the geographical location of the house, the construction standard etc. These quality factors need to be separated from real estate market risk, i.e. the price of risk which is independent of the quality factors. On therefore discriminates in the definition of transaction based real estate indices between price changes which are due to quality changes and pure real estate market price changes. To achieve this goal one constructs a **hedonic index.**

A hedonic index collects the price information of all transactions as follows. The contribution of the different price-sensitive characteristics of a residential property on the property price is estimated using multiple regression analysis. Quality factors are age, substance, distance to public transport, etc. ZWEX uses 24 characteristics. Each factor price is estimated using a regression. Once the factor prices are known, the heterogeneity problem of the sample of the houses is solved: The pure real estate price inflation can be separated from price variations which are due to the different characteristics. Let s_t^j be the price of the j-th residential property transaction at time t. We set

$$\ln s_t^j = \beta_0 + \delta_t \chi_t^j + \sum_{k=1}^{K} \beta_k^* \ln x^{jk} + \epsilon_t^j \ , \ t = T-40, T-39, \ldots, T-1, T \ , \tag{6.1}$$

with x^{jk} the k-th characteristic of object j, χ_t^j an indicator variable which takes the value 1 if a transaction took place in period t and zero if the object was sold in the period $t = 0$ and where the measurement takes place quarterly, i.e. one uses 40 quarters in the regression analysis. ϵ is the Gaussian noise term which is not correlated with the housing characteristics. The parameter δ measures the systematic growth of the real estate price level. The vector of coefficients β is assumed to be time independent, i.e. a price increase of an additional square-meter living space is assumed to be constant between the periods. Empirical tests show that the hedonic function and therefore also the index are better estimated if we use quarterly transactions from several periods and not only say the last periods. This indicates that real estate prices are strongly driven by a **drift** component compared to other assets classes where the **volatility** is the dominant price factor. The parameters are estimated using ordinary least square. The estimated hedonic function (6.1) can relate 85 percent of the house price differences between different objects to the considered characteristics. The standard error of the regression is at 14 percent. The hedonic method needs many data. Since the commercial real estate market is much more illiquid than the residential market one applies for the former one the discounted cash flow method with rents being the cash flows. Aggregating the DCF one obtains a so-called appraisal-based index.

We illustrate the method. Consider the factor 'distance from Zurich City Center'. Figure 6.4 shows a plot of house prices as a function of this distance.

We use the above method in a simplified form to estimate the different parameters and values. Write H_B for the price of land for construction in CHF we have:

$$H_B = \beta_0 + \beta_1 * \text{Dist} \ .$$

An OLS estimate implies

$$H_B = 1220 - 14.37 * \text{Dist} \ .$$

Therefore for each kilometer more distant to the center the square-meter price falls by 14.37 CHF. The R^2 is 0.7, i.e. at most 30 percent of the price variation is due to other factors.

Figure 6.5 shows the evolution of the transaction based ZWEX and Halifax Greater London Area indices. The index values are drawn without any modifications.

Although the index methodology are not exactly the same the main message is clear: While Zurich - and London area show a similar behavior in their property price development in the 80s up to the mid 90s of last century, the two areas faced a different evolution afterwards. The London area prices sharply increased compared to the Zurich prices. This reflects the increasing importance of London as the financial center. This increase was followed by a sharp correction during the financial crisis whereas Zurich showed still weakly increasing house prices. The interesting question is, how much between the price difference Zurich-London is due to a change in fundamentals and how much is bubble?

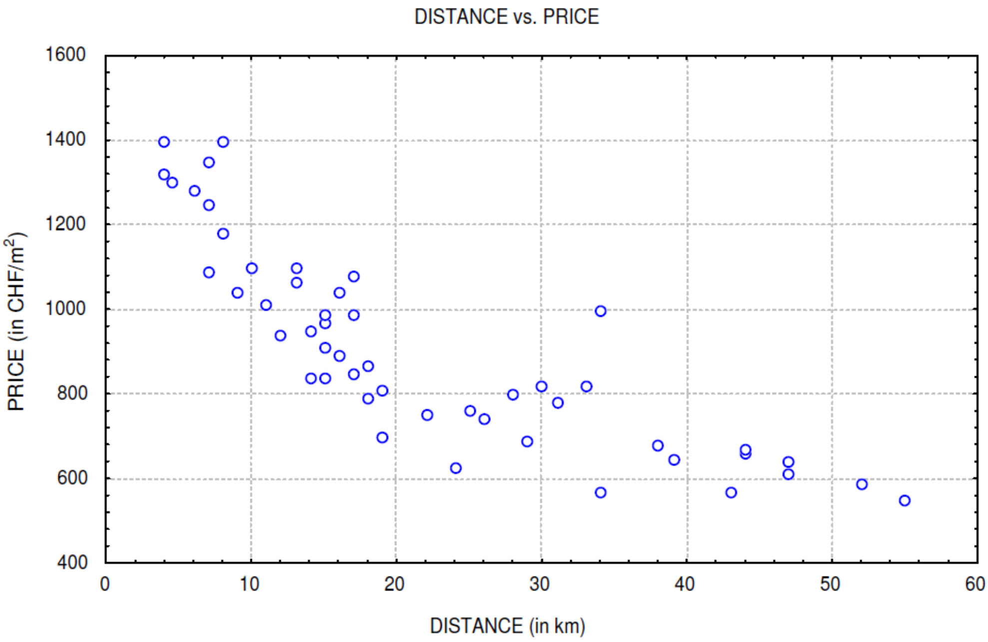


Figure 6.4: Scatter plot of house prices as a function of the distance to the Zurich City Center. *Source: Financial Engineering Immobilien, ZKB.*

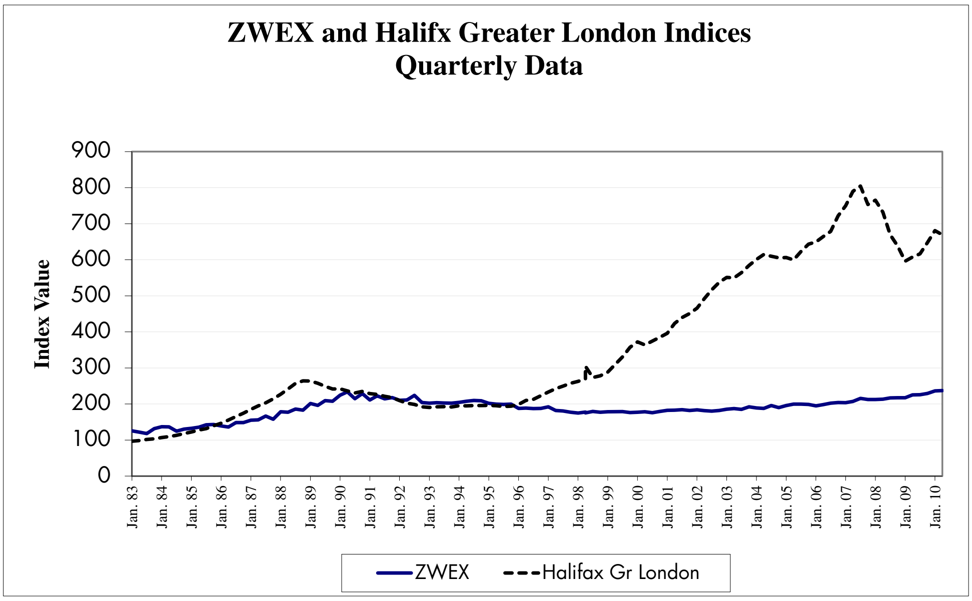


Figure 6.5: ZWEX and Halifax Greater London Area residential real estate price indices. Quarterly data.

6.1.3 Real Estate Derivatives and Mortgages

Having defined an underlying value one can **structure derivatives**. In principle, structuring of property derivative and pricing follows the same logic say as an equity derivative. But in practice several difference appear. First, the market for property derivative is mostly non-existing or at its best in a pre-mature state. Therefore, transactions take place in a back-to-back or, if the derivative firm has risk capacity and risk appetite, in a warehousing type. What makes the evolution of derivative markets for property difficult is the often observed homogeneity of beliefs - either there is a common belief that house price will raise or that they will fall. Compared to many other market a heterogeneous view on future property prices is less pronounced which makes back-to-back trading difficult and which drives the derivative house in their risk warehousing function fast towards its risk capacity limits. The most mature property derivatives market is the U.K. IPD derivatives market. At the end of 2008 some GBP 19.3 billion of swaps referenced IPD indices. Trading in IPD derivatives has decreased significantly by almost 80 percent from the 2008 figures. However, trading Eurex futures on IPD started in 2009 and is growing fast although it is still on a low absolute level of 250 million pound after Q1 2012. Property derivatives markets in France and Germany are also still very small.

An example of a property derivative are **index mortgages**. That is, a traditional mortgage with a protective put option on a residential property underlying index, such as the Halifax Index or ZWEX. The rationale to enter into such a contract for a home owner is to buy protection against falling house prices which will eventually force him to inject substantial fresh capital in the future. We assume:

- House price today of 1 Mio. CHF.
- Mortgage of 800'000 CHF., i.e. the LTV (loan-to-value) degree is capped at 80 percent of the house price.
- Therefore, the owner hat to inject 200'000 CHF equity to finance the house at the beginning.
- Suppose that the mortgage is of a fixed rate and fixed maturity type, say 5 years.

Assume that after 5 years house prices are down by 15 percent. The house is then worth 850'000 CHF. Eighty percent of this new value is 680'000 CHF. The bank is likely to reduce the mortgage notional to this value. This forces the owner to raise another CHF 120'000 of capital. The put option is designed to finance such a drop in house prices. That is, it reduces default risk of the home owner by smoothing his capital. The put option has the following payoff:

$$\max(K - S_T, 0)$$

after $T = 5y$ with S the value of the index and K the strike value. How do we price this option and how do the bank hedges the product? Since markets are incomplete pricing is a delicate matter and hedging is (almost) impossible. The traditional method in pricing is to use Black's model - although several assumptions are not satisfied. This model is used because of its often observed robustness: Although model assumptions are violated the use of more complicated models leads to prices which are not too far away from Black's model. The crucial input parameter in this model is the volatility of the underlying value. Mere observation of real estate indices shows that price dynamics of these indices is different than for more liquid, 'faster' markets such as equity, interest rates or commodities. Real estate indices show a kind of inertia. That is, if price raise or fall they gain a momentum which lowers volatility. This behavior is at least in line with the observed homogeneity of the beliefs. Using such a model, we obtain a fee between 50 and 80 bp per annum for the ZWEX. That is, if say interest rates are 2 percent for the mortgage, the client pays 2.5 to 2.8 percent p.a. and if a price decline of property follows the put option protects the clients' capital as explained above.

We consider the pricing of an **index-linked mortgage** with an embedded put option such that the notional is directly reduced by a potential negative index performance. We follow Syz et al. (2008). We are also interested in comparing the prices of index-linked mortgages with their traditional counterparts and in particular the valuation of collateralized and uncollateralized mortgages. To achieve this, we represent each loan as a linear combination of an unsecured loan, a credit derivative and, for the index-linked mortgage, a put option. This division follows the logic that an index-linked mortgage equals an unsecured loan plus a credit enhancement through a collateral, i.e. a credit derivative, plus an index put option. From a risk perspective, the risk factor in the unsecured loan is given by the default risk of the borrower, the risk factors of the credit derivative and of the put option are a combination of default risk and house price risk. We assume that the pricing of the put option in the index-linked mortgages and the rating of the borrowers are given.

We first consider a loan with a maturity of 5 years and fixed interest payments, secured by a collateral. The contract date is $t = 0$. The fixed maturity of the loan is T. The creditworthiness of the borrower is captured in an annual, constant default rate PD and $PD(0,T)$ is the cumulative probability of default in the period $[0,T]$, i.e. $PD(0,T) = 1 - (1 - PD)^T$. Therefore, we assume for simplicity that the creditworthiness of the obligor is stable over the whole lifetime of the contract. We consider the calculation of the default rate of a specific obligor and the valuation of a loan to the same obligor in two steps. First, we calculate the probability that the obligor defaults using several information sources: we use obligor, loan, and market-specific information to model the probability of default until time T using an internal based rating model. In the second step, we use the default probability to value all future cash flows of the financing contracts, i.e. the cash flows of the unsecured loans and the cash flows of a possible collateral for a situation in which the obligor might default at any time. Hence, the default probability enters in the discount factor of the cash flows.

Since default only takes place at maturity of the loan and that cash flows from the collateral are realized instantaneously at maturity. Therefore, the complicated and typically long-lasting recovery process following a default state shrinks to a single point, the maturity T. The annual refinancing rate with maturity t is the spot rate r_t. The discount factor for maturity t and for a compounding frequency of n is defined by $D_t = \left(1 + \frac{r_t}{n}\right)^{-nt}$. The notional loan amount is denoted by F. Coupon payments occur with a frequency n each year and *A(T)* is the cumulative amortization rate up to time T. The effective loan amount after T years of amortization is $\mathbf{F}_T = F(1 - A(T))$. We use Z to denote the periodic coupon payments including any amortization. We distinguish between three different interest payments: risk-free payments Z_{Rf}, payments from unsecured loans Z_u, and secured payments Z_s. rc denotes the expected recovery cost due to a default. These costs consist for example of remitted risk costs or lowered rates for the mortgage coupons. Moreover, there is a probability $1 - \gamma$ that the collateral of a defaulted debtor will not be realized, e.g., if the debtor leaves the default state after a financial reorganization. This probability varies between five and thirty percent in loan portfolios depending on the type of borrower, i.e. there are significant differences between private and corporate borrowers. Hence, the default state is not an absorbing state. If the collateral is realized, costs β follow, which we set proportional to the value of the collateral. These costs are incurred in trying to liquidate the collateral.

We write the value of the **unsecured loan** $B^{Z_u}(0,T)$ as the sum of a zero-coupon bond $B^0(0,T)$ and a coupon part $G^{Z_u}(0,T)$:

$$B^{Z_u}(0,T) = B^0(0,T) + G^{Z_u}(0,T) \tag{6.2}$$

with

$$B^0(0,T) = D_T[(1 - PD(0,T))\mathbf{F} - PD(0,T)rc] \tag{6.3a}$$

$$G^{Z_u}(0,T) = Z_u \sum_{t=1}^{nT} \left(1 - \frac{PD}{n}\right)^{nt} D_t \tag{6.3b}$$

The value of the **secured loan** is given by

$$B^{Z_s}(0,T) = B^0(0,T) + G^{Z_s}(0,T) + D_T PD(0,T)\gamma \mathbb{E}\left[\min\left(\mathbf{C}, \mathbf{F}\right)\right] \tag{6.4}$$

The second term represents the credit derivative, i.e. the credit enhancement. $\mathbf{C}$ is the effective collateral which reads $\mathbf{C} = \min\left(N, \max\left((1-\beta)H_T + a_C - v, 0\right)\right)$ with v the value of any senior debts, H the value of the property object, β the average estimated costs in the recovery workout, N the value of the borrowers note and a_C the value of collateral other than property. If the collateral is realized, the bank receives the smaller of $\mathbf{C}$ and $\mathbf{F}$, i.e. $\min(\mathbf{C}, \mathbf{F})$. Using $\min\left(\mathbf{C}, \mathbf{F}\right) = \mathbf{F} - \max\left(\mathbf{F} - \mathbf{C}, 0\right)$, the present value of the loan reads

$$\begin{aligned}B^{Z_s}(0,T) &= B^0(0,T) + G^{Z_s}(0,T) \\ &+ \underbrace{D_T \gamma PD(0,T) \left[\mathbf{F} - \exp(r_T T)\mathrm{Put}\left(\mathbf{F}, T, r, \sigma_{\mathbf{C}}, \mu_{\mathbf{C}}\right)\right]}_{=:CD(0,T)} .\end{aligned} \tag{6.5}$$

with $CD(0,T)$ the credit derivative written on the collateral, $\mathrm{Put}(\cdot)$ a European put option, $\sigma_{\mathbf{C}}$ the volatility of the effective collateral, and $\mu_{\mathbf{C}}$ the annual growth rate of the effective collateral. The value of this put option can be determined following the approach of Schiller and Weiss (1994). Their formula is similar to the Black-Scholes formula except that the underlying asset grows at a rate other than the risk-free rate. This difference is due to the fact that the collateral is not necessarily a liquid asset, hence it can not be effectively used to hedge the option. Assuming that the effective collateral is log-normally distributed the put option is worth

$$\mathrm{Put} = \exp(-rT)\left[\mathbf{F}N(d_1) - \mathbf{C}\exp\left(\mu_{\mathbf{C}} + \frac{1}{2}\sigma_{\mathbf{C}}^2\right)N(d_2)\right] \tag{6.6}$$

where $d_1 = \frac{ln(\mathbf{F}/\mathbf{C}) - \mu_{\mathbf{C}} T}{\sigma_{\mathbf{C}}\sqrt{T}}$, $d_2 = d_1 - \sigma_{\mathbf{C}}\sqrt{T}$ and N is the standard normal distribution function.

The no-arbitrage argument implies that the rates for all types of financing - risk less, unsecured, and secured - are such that their present values are the same. In other words, the equation

$$F = B^{Z_{rf}}(0,T) = B^{Z_u}(0,T) = B^{Z_s}(0,T) \tag{6.7}$$

determines the coupons, which can be calculated analytically. For instance, Z_u follows from (6.2) as

$$Z_u = \frac{F - D_T[(1 - PD(0,T))\mathbf{F} - PD(0,T)rc]}{\sum_{t=1}^{nT}\left(1 - \frac{PD}{n}\right)^{nt} D_t} . \tag{6.8}$$

Z_s is calculated in the same way except that, due to the put option, there will be an additional term in the numerator of (6.8). The model is applied to the pricing of the loan contracts for 8 different rating classes, with class 8 the defaulted class, see Figure 6.6.

The upper panel shows that the lower the creditworthiness of a borrower, the more expensive the terms for the unsecured loan. Second, the lower the creditworthiness of the borrower, the more the bank (as protection buyer) is willing to pay for collateral. shows that uniform in the borrower's creditworthiness the terms of the index-linked mortgage are more expensive than for the classical one. But one observes that the bank's willingness to pay for the credit enhancement depends on the creditworthiness of the borrower: the higher the creditworthiness, the lower the compensation for the credit enhancement. The difference in the final terms between the two mortgages for the rating class 7 is 0.94%

Rating	Loan Principal	Estimated Value of Property	Value of Borrower's Note	Unsecured Loan	Collateral Enhancement	Terms of Mortgage
7	500	625	500	14.21%	8.54%	**5.67%**
6	500	625	500	8.20%	3.68%	**4.52%**
5	500	625	500	5.55%	1.43%	**4.12%**
4	500	625	500	4.43%	0.47%	**3.96%**
3	500	625	500	4.25%	0.32%	**3.94%**
2	500	625	500	4.04%	0.13%	**3.91%**
1	500	625	500	3.95%	0.05%	**3.90%**

Standard Mortgage				Index-Linked Mortgage		
Rating	Unsecured Loan	Collateral Enhancement	Terms of Mortgage	Index Put	Credit Enhancement	Index-Linked Mortgage
7	14.21%	8.54%	**5.67%**	1.09%	0.15%	**6.61%**
6	8.20%	3.68%	**4.52%**	0.95%	0.07%	**5.41%**
5	5.55%	1.43%	**4.12%**	0.88%	0.03%	**4.97%**
4	4.43%	0.47%	**3.96%**	0.81%	0.01%	**4.76%**
3	4.25%	0.32%	**3.94%**	0.72%	0.01%	**4.66%**
2	4.04%	0.13%	**3.91%**	0.71%	0.00%	**4.62%**
1	3.95%	0.05%	**3.90%**	0.71%	0.00%	**4.61%**

Figure 6.6: Upper Panel: Pricing of interest rate type mortgages. The final price consists of the unsecured loan and the collateral enhancement for the bank as a protection buyer. We have chosen the following parameters: β is CHF 20′000 / CHF 625′000 = 3.2 percent, probability of recovery $1-\gamma$ is 30%, volatility of the collateral σ_C is 6% and the growth rate of the collateral μ_C is 1.5%. For the risk free interest rates we used CHF Swap Rates as of October 25, 2005 and Moody's default statistics to obtain the credit risk dependent discount factors. Lower Panel: Pricing of interest rate type mortgages and index-linked mortgages. Index-linked mortgages are always more expensive than interest rate type mortgages. But the lower a homeowner's creditworthiness, the less are the relative additional costs for the put option compared to the interest rate case. Data are the same. For the best two rating categories 1 and 2, the put premium is approximately 0.7% p.a. for a 5 year term, correlation between the index and the collateral is estimated 75 percent and the strike of the index put is set equal to the initial value of the index, i.e. at 100 percent. *Source: Syz and Vanini (2008).*

whereas the put premium is 1.09% . This result is due to the following ideal assumptions: the loan makes 80% of the estimated house price and the borrower's note equals the value of the mortgage's principal. However, if the loan policy is tight, index-linked mortgages reduce refinancing risk essentially.

How does a bank hedges such a put option, i.e. how is a portfolio of instruments set up such that the portfolio value matches the value of the put option and where the price of such a portfolio is given by the paid premia of 50 to 80 bp per annum? Given the pre-mature state of the real estate market, there is no simple standard hedging by going long or short the respective products. In fact, the bank can take a bet that real estate price will raise and not fall and keeping the position open. This is a dangerous game which requires an appropriate risk capacity and risk capital of the bank. A proxy can be constructed cross business units as follows. The put option is issued by the trading unit; the mortgage is a retail banking unit business. Therefore, if the put option is in the money, house prices are falling. The bank then faces the risk that the premia paid for the put option are less than the value of the option after 5 years, i.e. the bank faces a trading loss. The retail banking unit on the other hand side faces less counter party risk due to the protective put option. That is, everything else the same, the margin gained by the retail banking is higher due to lower risk costs. Since risk costs in Switzerland for residential house financing are low - 50 bp per annum is yet a large figure - the value of the option premia of say 50 bp p.a. and the lower counter party risk premia of the same size add to a 5 percent figure over five year. Therefore, if the drop in house prices is not larger than 5 percent, the integrated view across the business units provides a hedge. If house prices drop more, the hedge becomes only approximate and net losses follow. Since a 5 percent decrease in house price over five years is only a moderate amount, this proxy hedge is not effective to cover more extreme events. In Switzerland, at the beginning of the 90s house prices in the Zurich area dropped by 14 percent. In some districts of Miami after the financial crisis 2008 prices were down by 30 percent or more.

There were in total 2 transactions of these protected mortgages since the issuance of the products in 2006. A major reason is that although each house is unique, beliefs about house price evolution is homogeneous. One can think hardly about a different market where people are either all bullish or all bearish. It is an open issue to show why this is the case.

We conclude this section with an innovation of an investment in a retail structured product. The product is defined by:

- Underlying value: ZWEX
- Maturity: 5y
- Currency: CHF
- Issuer: AAA bank

- Size of issuance CHF 50 Mio.
- Issuance price: 100 percent.
- Payoff: $100 + \max(0, \min(Z_T, 5.5\%))$ with Z_T the ZWEX performance over 3y.

This product therefore pays pack in the worst market case 100 percent of the invested capital at maturity. The return is capped at 5.5 percent. To price the option payoff $\max(0, \min(Z_T, 5.5\%))$ one should apply methods of incomplete markets. But in practice one use the robust Black model. Although the models assumption are violated for property derivatives on the ZWEX one often uses this model. First, it is not clear whether violating some assumptions in a model is more severe than using a more sophisticated model were many assumptions are finally needed to calibrate the larger model. Second, finance is not a natural science. That is, if many market participants apply a model which in principle could not be applied but if the usage of this model is common knowledge then demand and supply can emerge and match in such a setting.

6.1.4 Pricing of Property Derivatives

Prices of real estate indices show an inertia property: Van Bragt et al. (2009) state: *' ... autocorrelation can occur in appraisal-based indices because appraisers slowly update past prices with new market information. Transaction-based indices can also exhibit a positive autocorrelation because private real estate markets are less informationally efficient than public securities markets. As a result, the price discovery and information aggregation functions of the private real estate market are less effective. This can cause noisy prices and inertia in asset values (and returns).'* See also Fabozzi et a. 2009, Geltner and Fisher (2007) for other work on pricing.

We consider two approaches. The model of van Bragt et al. (2009) and the approach of Syz and Vanini (2011). Van Bragt et al. (2009) state: *' ... Geltner and Fisher (2007) note that a 2006 survey of (potential) market participants identified a lack of confidence in how real estate derivatives should be priced. They also note that this concern is understandable, since the underlying asset cannot be traded in a frictionless market. This makes it impossible to use classic pricing formulas for derivatives (such as the relationship between spot and forward prices), since these formulas only apply under strict no-arbitrage assumptions. Geltner and Fischer (2007) argue, however, that the valuation of real estate derivatives is still possible using equilibrium pricing rules, provided that the dynamic behavior of the underlying real estate index is properly taken into account.*'

Bragt et al. (2009) take the next step by proposing a quantitative risk-neutral valuation model which can be used for actual pricing purposes. Their model is a discrete time model for the observed real estate index in combination with continuous-time models for the efficient market process of real estate and for interest rates. They first consider the real-world price process. They assume that the following first order dynamics representing auto-regression of a real estate index holds:

$$A_t = kS_t + (1-k)(1+\pi)A_{t-1} \tag{6.9}$$

with A the current price, S the unobservable true market price, $0 \leq k \leq 1$ a constant an π the expected (annual) return of the index which accrues the value A_{t-1}. Solving this first order difference equation shows that it is equivalent to an exponentially weighted moving average. The solution delivers the index return r^A based on A and the unobserved market return based r^S on S. They then assume that the underlying market returns follow a random walk process with drift, i.e. $r^S = \pi + \epsilon_t$ is the market return with the normally-distributed, serially-uncorrelated noise term with zero mean and variance σ_ϵ^2. The model can be extended by (i) considering more than one leg and by (ii) using seasonality. The authors then consider risk-neutral pricing. In risk-neutral world all individuals expect to earn on all securities a return equal to the risk-free rate. This simplifies the valuation of options since the option payoffs can simply be discounted along the path of the short rate for each scenario. The authors choose a short rate interest rate model - the Hull-White model. This model is used to price the expected return π. That is we have the same formula for the index updating (6.9) but π is equal to the value of a money market account in one period times a correction for the direct return g associated with real estate investments:

$$\pi(t) = \exp\left(\int_{t-1}^{t} r(s)ds\right) e^{-g}$$

with $r(s)$ the nominal short rate satisfying the Hull-White model dynamics. The authors considers whether in their model arbitrage is excluded. That for they have to verify that the discounted tradeable asset A is a martingale under an appropriate measure. The analysis shows that:

- If all direct returns g are reinvested in the index. Then the price index becomes a total return index. For such an index the martingale property holds for the efficient market process S with numeraire the money market account. If returns are paid out, the martingale property does not hold: A price index is thus not a tradable asset if direct returns are paid out.

- Consider the realization of the real state index A normalized by the money market account. Only when $k = 1$, i.e. the $A_t = S_t$, the martingale property holds. Arbitrage opportunities thus exist in case of a complete market when trading an autocorrelated real estate index, i.e. $k < 1$. The reverse also holds: the index value may well be different from the efficient market price, but active trading in the index is not possible in this case: otherwise arbitrageurs would quickly force the index value toward the efficient market price.

We consider Syz et al. (2011). Their starting point is that market frictions inhibit perfect replication of property derivatives and they define the property spread as a price measure in the incomplete real estate market. Syz et al. (2011): *'In a swap that pays the total return of a property index, the rate that balances the swap can deviate significantly from LIBOR. We call this difference between the property swap rate and LIBOR the* ***property spread****, quoted on an annual basis. The swap payer pays property performance and in return gets LIBOR plus the property spread. If the spread is negative and its absolute value exceeds LIBOR, the swap payer pays on the interest leg of the swap but expects to receive the negative performance of the property leg. In the UK, property derivatives were traded at a substantially positive spread until the end of 2006. However, the spread fell in 2007 and quickly turned negative. Quotes obtained from market participants who trade swaps on property indexes differ considerably from prices computed using models based on arbitrage arguments. In contrast to property returns, equity returns are swapped against LIBOR without a spread. The reason is that a no-arbitrage argument is sufficient to price equity derivatives. A trader can sell short equities at virtually no transaction cost and invest the proceeds in an instrument returning LIBOR. It would thus be a free lunch to receive a rate higher than LIBOR.'*

However, a no-arbitrage argument alone is not sufficient to price property derivatives because the underlying market exhibits frictions. The index and its components cannot be traded continuously and instantly at the prevailing spot price without transaction costs. This leads to a property spread. Observed property spreads vary with the maturity of the swap. The cause of the shape of the term structures of property spreads is not obvious. As liquid and cost efficient instruments, property derivatives are beneficial to both investors and hedgers. Given the significant transaction cost advantages, it is clear that market participants looking for a short-to-medium term property exposure or hedge can benefit from the use of property derivatives. For long-term investment horizons, the impact of one-one transaction costs is less significant, making a physical purchase or sale a viable alternative to a property swap. Thus the short end of the term structure of property spreads is expected to be more volatile than the long end. The common explanation of a cash-and-carry arbitrage[1] fails since it implies an inverse spread curve against maturity. Such a qualitative shape was observed in bullish markets 2007 but not in bear markets 2008, see Figure 6.7.

Since we observe severe frictions that inhibit perfect replication all explanations such as the cash and carry arbitrage will always be of limited value to explain pricing behavior. The property spread exists because property derivatives cannot be perfectly replicated by trading actual property. These frictions define arbitrage free price bounds for the property spread. The authors consider (i) transaction costs, (ii) transaction time and (iii) short sale constraints. For the US, aggregate agent's fees for housing transactions range from 3% to 6% as in DiPasquale and Wheaton (1996), in OECD countries, round trip transaction costs are generally estimated to range from 6% to 12% as in Quigley (2002). However, technologies such as online marketplaces have already begun to reduce

[1] A cash-and-carry arbitrage occurs when a trader borrows money, buys the goods today for cash and carries the goods to the expiration of the futures contract. Then, delivers the commodity against a futures contract and pays off the loan. Any profit from this strategy would be an arbitrage profit

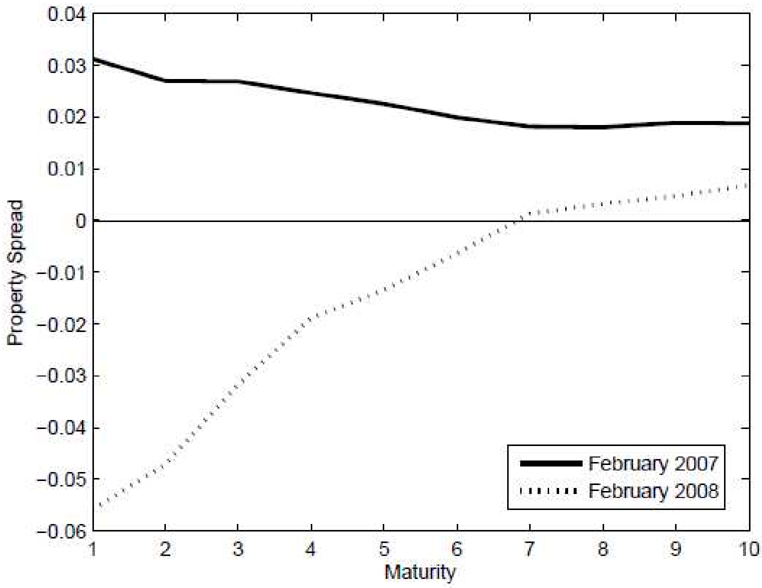


Figure 6.7: The term structure of property spreads. The lines show the property spreads of Halifax HPI contracts against maturity, based on ask prices, in February 2007 and one year later. *Source: Syz and Vanini (2011).*

some of these costs for homeowners. It is not only costly but also time consuming to trade physical property (transaction time). It is a fact that the property market is not quite transparent. In many regions and for some sub-markets, very few comparable transactions can be observed to indicate a price level. Consequently, uncertainty about demand and price for an individual object is high. Transaction time reflects market illiquidity. In the UK, it takes on average three months to find a counter party and another three months to finalize a transaction. Derivatives in contrast can be traded almost instantly. To be able to sell an asset short one must borrow it. The borrower pays a lending fee to the lender. Because of the short sale constraint, arbitrageurs can only refrain from buying overpriced assets but cannot exploit mispricings. These frictions define an incomplete market - the risk neutral probability is not unique, or bid and ask prices are not unique but define a possibly infinitely bounded interval of arbitrage free prices are the same statements. Given the frictions, bounds of arbitrage free prices follow rather than one single arbitrage free price for property derivatives. The arbitrage free price bounds are a function of the price of the underlying instrument and of market frictions. For any given property spread, there is an upper arbitrage free price bound $\bar{p}$ and a lower arbitrage free price bound $\underline{p}$. The upper bound is the maximum spread an investor is willing to pay for a derivative instead of buying actual property and is only affected by buyer and seller transaction costs and by transaction time. If the property spread lies above the upper arbitrage free price bound, it is more attractive to buy actual property than to buy derivatives. Unlike the upper bound, the lower bound also reflects the value of the short sale constraint. The authors determine the upper and lower bounds in term of the friction parameters, i.e. the cost factors of the three friction types. They then estimate these cost parameters. Fig. 6.8 plots the historical trajectory of the Halifax HPI and the arbitrage free price bounds for its forward prices, using the obtained values for the market frictions.

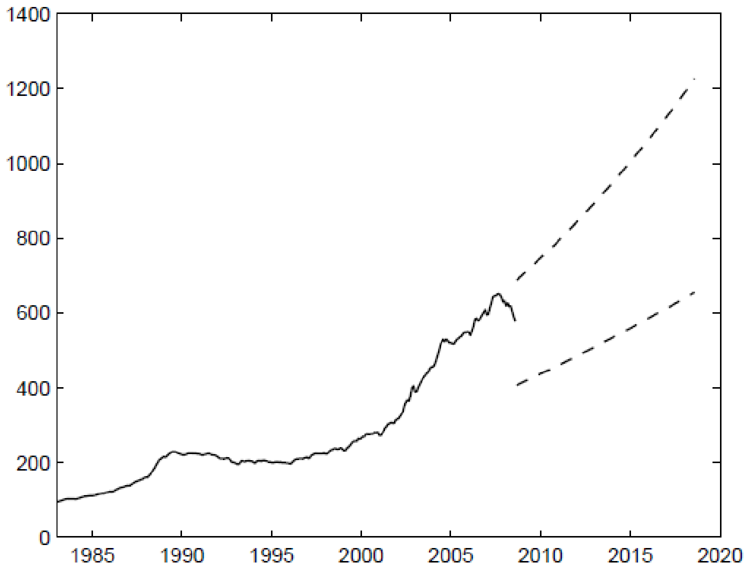


Figure 6.8: Arbitrage free price bounds. The figure plots historical levels of the Halifax HPI (1983=100) and the arbitrage free price bounds for forward prices of Halifax HPI index contracts (dashed lines). *Source: Syz and Vanini (2011).*

6.2 Green Banking

The climate change is one of the major threats and opportunities in this century. We first provide an overview about some key facts of the climate change including its potential impact on society, bio diversity and the economy. Different examples are given where the interplay of financial innovation and technological progress leads to ecological and economic meaningful solutions. Summarizing, technologies are developed such that there are alternative solution to the government's approach of command & control.

Although there are still some controversies, data show that humankind is facing some challenges. Energy demand will further increase due to population growth, progressing industrialization and increasing wealthiness. This will without any countermeasures increase CO2 emissions. But climate change requires the opposite, i.e. drastic reduction of CO2.

We show in this section how solutions based on technological and financial innovation can create situation, where it pays economically to reduce CO2 emission and to invest in energy efficiency. The results depend on the following facts:

- There is enough clean energy, i.e. energy which can be used to replace CO2 emitting energy.
- The technology to transport energy efficiently exists.
- There exist financial market solution which match investor's demand for sustainable investment with the demand for energy project finance.

6.2.1 DESERTEC

To highlight the first point, we consider the DESERTEC concept. The bottom line of the project is that a **300 times 300 kilometer** thermal solar energy plant in a desert is sufficient to **generate enough energy** to cover the world wide electricity demand. DESERTEC includes energy security and climate protection as well as drinking water production, socio-economic development, security policy and international cooperation. DESERTEC is a German foundation. So far, the DESERTEC Foundation has already made significant progress in the Mediterranean region. We comment on some risks and facts of this project.

- **Political risk.** One could argue that some countries such as Algeria, Libya, Saudi Arabia, the VAR state given their natural oil resources have no interest in a solar energy project. This is not the case due to the following reasons. First, the solar energy project will be beneficial for employment and job creation in these countries to a far larger extent. Second, due to the excess solar energy these states will be able to create new farming land. What indeed is a risk is the comeback of colonization emotions and feelings. Why should these countries help European or general Western countries? This is a delicate issue which asks for prudential and earnest negotiation. Another risk factor is political instability. The events starting in 2011 demonstrate that this risk is there. But since the project intends to generate a win-win situation for the Sahara related countries and the others this risk is best minimized by considering the socio-economic development seriously. A political risk is prominent in Europe - where will the energy enter the continent? The Northern European states do not want become dependent on a single point of entry in the Mediterranean region.

- **Technology and financing risk.** Since thermal solar energy and not photovoltaic defines the technology the concern that sand and sand storms can heavily damage the energy production is not relevant. Also losses in the transport of electricity are with today' energy standard no longer relevant. The status of the energy infrastructure together with the financial weakness of many European countries rises a significant risk since new infrastructure needs to be build up and existing one has to be renewed. At this point it becomes evident that financial innovation is necessary to bridge potential financing gaps.

DESERTEC is an initiative in the context of climate change. Figure 6.9 shows the impact of potential climate changes on different dimension of humanity and the ecosystem.

We first note that the climate is **already changing**. The increase in CO2 over the last 100 years has lead to a measurable change in climate. The question is therefore not if this will happen but to what extent. The data from researchers show that climatic change less related to risk but more to uncertainty: There is lack of knowledge about the speed, the irreversibility, possible feedback effects and some hidden non-linearities. The impact of the worst case scenarios on **GDP** forecasts a drop between 5 and 20 percent. Estimates of Credit Suisse state that investment flows of at least USD 700–850 billion per annum will be required over the next decade to limit warming to 2 degrees Celsius and to adapt to the effects of changes that cannot be avoided. This is a conservative estimate. Other estimates range up to USD 2'000 billion. The majority of the required capital investment is concentrated in low carbon energy, energy efficiency, and low carbon transport infrastructure. Low carbon energy is primarily linked to investment in renewables, electricity infrastructure like grids and transmission and storage. The opportunity is concentrated in China, the US and the EU27. They represent nearly 60 percent of the mitigation cost. Figure 6.10 shows the distribution of the investments necessary to achieve the 2 percent pathways. The matrix has the dimensions geographical location and area of investment. The authors WWF and Credit Suisse (2011) state different type of barriers affecting the current decarbonization efforts. They state that *'low carbon investment is directly linked to government intervention, and still falls short of what is needed.' WWF and Credit Suisse, 2011.* Since regulatory mechanisms do not exist yet which price the externalities of carbon emissions technical and financial barriers can exist. Simply, the economics of low carbon projects are often less attractive than those of their high carbon alternatives. Structural barriers include network effects (consumers

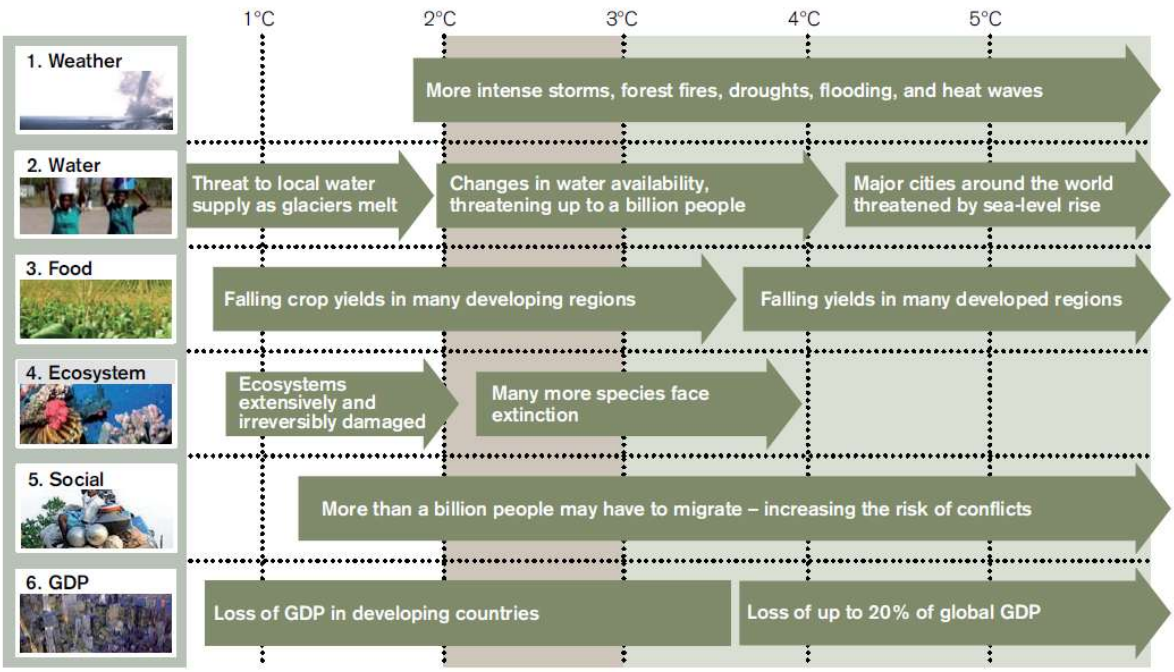
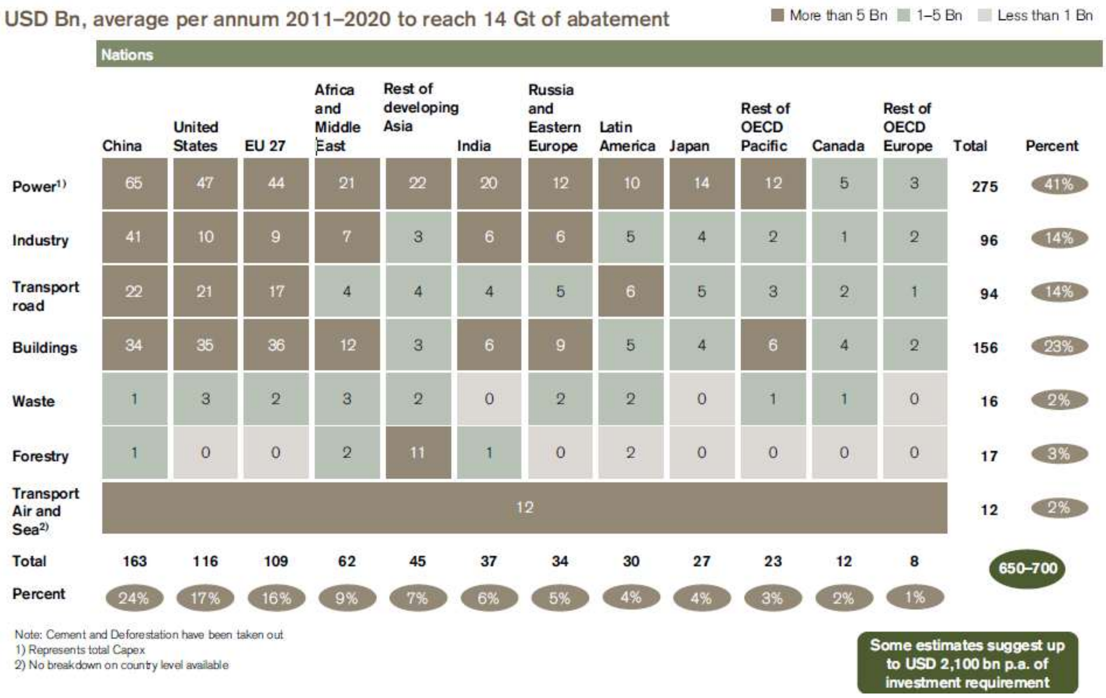


Figure 6.9: Impact of potential climate change. *Sources: Stern Review, IPCC, 4th Assessment Report, Climate Change 2007, WWF and Credit Suisse 2011.*

will not buy electric cars unless there are workable and available charging solutions, but no private investor will build a charging network unless there is sufficient demand from electric vehicle users) , agency problems (the party making the low carbon investment is under existing structures often not the one which will benefit from the savings) or the status quo bias (strong bias towards maintaining the status quo instead of making changes). The last point is often called 'change resistance' behavior.


USD Bn, average per annum 2011–2020 to reach 14 Gt of abatement

More than 5 Bn | 1–5 Bn | Less than 1 Bn

Nations

	China	United States	EU 27	Africa and Middle East	Rest of developing Asia	India	Russia and Eastern Europe	Latin America	Japan	Rest of OECD Pacific	Canada	Rest of OECD Europe	Total	Percent
Power[1)]	65	47	44	21	22	20	12	10	14	12	5	3	275	41%
Industry	41	10	9	7	3	6	6	5	4	2	1	2	96	14%
Transport road	22	21	17	4	4	4	5	6	5	3	2	1	94	14%
Buildings	34	35	36	12	3	6	9	5	4	6	4	2	156	23%
Waste	1	3	2	3	2	0	2	2	0	1	1	0	16	2%
Forestry	1	0	0	2	11	1	0	2	0	0	0	0	17	3%
Transport Air and Sea[2)]	12												12	2%
Total	163	116	109	62	45	37	34	30	27	23	12	8	650–700	
Percent	24%	17%	16%	9%	7%	6%	5%	4%	4%	3%	2%	1%		

Note: Cement and Deforestation have been taken out
1) Represents total Capex
2) No breakdown on country level available

Some estimates suggest up to USD 2,100 bn p.a. of investment requirement

Figure 6.10: Annual investment required to achieve 2 degrees Celsius pathway is USD 650–700 bn. *Sources:Credit Suisse/WWF analysis based (2011) on McKinsey's Climate Desk tool.*

When it comes to financial innovation, several directions can be considered. First, banks need to understand the new types of risk. This together with the demand for such risk allow them then to define new products and services (risk structuring and risk transfer). Banks can next use their experience to raise new capital. This can follow traditional paths such as bonds or direct placements. Other forms can be more in the form of Structured Finance or Securitization. This form of risk structuring and risk transfer came heavily under fire in the last financial crisis. But it could well be that the principle mechanism of these innovation will be very fruitful in the future to raise capital.

6.2.2 Failures and Examples

Although the above description seems optimistic large projects may fail. The Lisbon Strategy, adopted in 2000, which it is generally agreed largely failed in its grand attempt to turn the EU into 'the world's most dynamic knowledge-based economy by 2010'. The approach was founded on three pillars: An **economic pillar** to prepare the ground for the transition to a competitive, dynamic, knowledge-based economy. A **social pillar** designed to modernize the European social model by investing in human resources and combating social exclusion. An **environmental pillar** recognizing that economic growth must be decoupled from the use of natural resources. The Lisbon process established a lengthy list of targets, most of which fell out of the control of EU institutions. In 2005, the strategy's failures were already very visible: an overly complex structure with multiple goals and actions, an unclear division of responsibilities and tasks and a lack of political engagement from the member states. A report led to the 2005 re-launch of the strategy, focusing almost exclusively on growth and jobs, and with a slight adaptation of the governance model to enhance the partnership between the European Commission and the member states. The 2007 financial and economic crisis was then the final blow to the Strategy.

Meanwhile, the EU faces serious competitive threat of the upcoming economic powerhouses such as China and India. Furthermore, the environmental agenda has also moved on. Under the imperatives of climate change, the new thinking had to take a broader view and bring together the economic, social and environmental agendas of the EU in a more structured and coherent way. At the Spring Summit 2010 EU leaders endorsed the European Commission's proposal for a Europe 2020 strategy, replacing the heavily criticized Lisbon Strategy, and this Europe 2020 Strategy is now the centrepiece of the Commission's mandate. This new strategy puts knowledge, innovation and green growth at the heart of the EU's blueprint for competitiveness and proposes tighter monitoring of national reform programmes, one of the greatest weaknesses of the Lisbon Strategy. The basic proposal is to continue to promote EU growth based on knowledge and innovation, alongside high employment and social cohesion, while promoting a sustainable perspective (both in competitive and environmental terms).

Another example concerns water pollution in Switzerland in the 60s of last century. Defining incentives and providing finance start up help by the Swiss government a new industry emerged (clarification plants) and treatment of farming land was changed. After some decades water from Swiss lakes or rivers is potable.

Other examples of market based solutions are from the US experience with curtailing acid rain via the sulfur dioxide allowance market to the implementation of the Clean Water Act. Market-based solutions have proven consistently **more effective** in protecting the environment than government regulation alone. Project financing, public/private partnerships, and tradable permits have come to supplement or replace conventional command-and-control regulation and purely tax-based instruments.

This approach can minimize the aggregate costs of achieving environmental targets while providing dynamic incentives for the adoption and diffusion of greener technologies. Market-based financial and public policy instruments emerged in the 1980s and have steadily gained momentum. In fact, the United Nations Environment Programme has launched a Finance Initiative as a formal mechanism for mobilizing the financial sector to take a more active role in protecting the environment. For many years, the US financed environmental improvements such as sewage treatment plants with substantial one-time grants made by the federal government to localities. But as the demand for funding grew, policymakers and capital market experts began to think creatively about how to leverage finances: to generate, for example, USD 2 million of pooled capital from a one-time USD 500,000 grant. In 1987 the Federal Clean Water Act replaced its aging grants program with state revolving funds. Under this model, each state applies for a federal capital grant, which requires a 20% local match; this pool of funding is then supplemented with capital market investment and possibly 'seed money' from philanthropies. The states then make low-interest loans to local municipalities and organizations, which repay the loans from project revenue and local taxes. States may provide loans to communities, individuals, nonprofit organizations, and commercial enterprizes. Their repayments recapitalize the state fund, creating a sustainable resource for funding. This model maximizes the impact and longevity of government grants.

The most practical solution for building a greener economy is to correct faulty pricing by making consumers and firms pay for the environmental damage they cause. Once these **negative externalities are internalized**, they will be incorporated in the prices of goods and services, creating real incentives for the creation and adoption of clean technologies. One of the most compelling examples of using these principles to fix broken markets is that of cap-and-trade pollution markets. In such markets, the cap (or maximum amount of total pollution allowed) is usually set by government. Businesses, factory plants, and other entities are given or sold permits to emit some portion of the region's total amount. If an organization emits less than its allotment, it can sell or trade its unused permits to other businesses that have exceeded their limits. Entities can trade permits directly with each other, through brokers, or in organized markets.

A further example are **Debt-for-Nature Swaps (DNS)**, i.e. to relate a country's external debt and its ability to protect bio-diversity. Debt-for-nature Swaps involve the purchase of a developing country's debt at a discounted value in the secondary debt market and cancelling the debt in return for environment-related action on the part of the debtor nation. By relieving the foreign debt burden carried by developing nations, it is possible to secure their commitment to invest in local conservation projects and save ecosystems. The concept of debt-for-nature swaps was first conceived by Thomas Lovejoy of the World Wildlife Fund in 1984 as an opportunity to deal with the problems of developing-nation indebtedness and its consequent deleterious effect on the environment. Lovejoy called for building an explicit link between debt relief and environmental protection.

The DNS is a system in which international NGOs take the lead in swapping foreign debt accumulated in developing country's for nature protection project in the country's own currency: International NGOs will purchase some of developing countries' debts from private banking institutions and invest their own currency so that those countries may work for nature protection in return for a lapse of the debts.

The first swap happened in 1987 between the Bolivian government and Conservation International (CI). The government agreed to protect 4 million acres of forest and grassland in the Beni Biosphere Reserve with maximum legal protection. The deal took eight months to complete, because of the lack of open participation by organizations in Bolivia and some misplaced perceptions about DNS and how it would work. Many Bolivians originally believed that land from Bolivia was being transferred to Conservation International. Another important note to make is although Bolivian government could have bought back its own debt on the secondary market, the money would not have stayed within the country (for environmental projects) without the DNS. Debt-for-Nature swaps gained momentum and support when President George Bush included DNS in his 'Enterprize for the Americas' initiative. Several countries have participated in at least one debt-for-nature swap. These countries include: Madagascar, Zambia, Bolivia, Costa Rica, the Dominican Republic, Poland, Nigeria, the Philippines, Brazil, Panama, and Cameroon. Three international conservation organizations, all of which are based in the U.S., have been the most active: Conservation International (CI), The Nature Conservancy (TNC), and the World Wildlife Fund (WWF). Measured by either the face value of the debt or by the amount of funds to the conservation organizations, Costa Rica, Ecuador, the Philippines, and Madagascar, have been most heavily involved the countries.

Some critics of DNS suggest that the debts in many 3rd World countries are so large that these buyback schemes are of very little value. By 1992, 17 countries had signed DNS agreements and while donors spent USD 16 million to buy back USD 100 million in debts, the nominal reduction barely touched these country's debt burdens. When Bolivia spent USD 34 million to buy back USD 308 million in bonds in 1988, the price of the remaining bonds rose from 6 cents on the dollar to 11 cents on the dollar. As a result, the real value of the outstanding debt declined from USD 40.2 million dollars (USD 670 million at 6 cents on the dollar) to USD 39.8 million (362 million at 11 cents on the dollar).

6.2.3 Energy Contracting and Structured Finance

We consider **energy contracting solutions**. Such contracts are defined between the following parties:

- Energy efficiency searching institution. An institution - public entity, a corporate - in our case wants to reduce energy costs in an existing building or a new project. To be specific we consider a large city administration.
- Energy solution provider. A corporate provides the technology to realize the energy

cost gains. The energy solution should lead to a substantial reduction in energy costs.

- Financial solution provider. A bank offers different possibilities to finance the project. The financial solution should reflect the particular financial and political needs of the city administration.

As an overview the following figures hold as rough rules in case buildings are made energy efficient. The data are from Siemens AG.

Type of Optimization	Energy Saving	Amortization Period
Measure & Visualize	$\sim 10\%$	$\sim 1-2$y
Optimizing Operations	$\sim 10-25\%$	$\sim 2-3$y
Building Services Eengineering	$\sim 25-35\%$	$\sim 3-7$y
Renew	$\sim 30-45\%$	$\sim 6-12$y

Table 6.1: Rough figures for building energy efficiency measures.

Measure and Visualize means that a firm makes transparent its energy consumption at well-chosen location within the firm. Elevators are often used since most people in an elevator search for a fix point to focus on or the entrance lobby is also well suited. It has by now been reported in several studies that simple transparency or monitoring without any other actions leads to an approximate energy reduction of about 10 percent. It seems that such a transparency changes behavior of some employees leading to this reduction.

When it comes to financing the project a major requirement is that the project also makes economic sense. That is, we require **Gain** > 0**.** The gain is a sum of investment costs I and the savings of energy costs over time. Saving of the energy costs has four risk sources:

- Investment risk. The amount $I = \bar{I} + dI$ is equal to the expected costs $\bar{I}$ and possible deviations dI.

- Volume risk. I.e. the amount of saved energy c_t is given by

$$c_t = \bar{c} + dc_t$$

  with $\bar{c}$ the expected amount of saved energy once the project is finished and dc_t the risk of deviation from the expectation.

- Energy price risk. I.e. the price p_t of saved energy (oil, electricity, a mixture of them) is equal to

$$p_t = \bar{p} + dp_t$$

  with $\bar{p}$ the forward/futures prices and dp the deviation risk from the forward prices.

- The last risk is counter party risk of the energy solution user - here the city administration. Depending on type of financing the project the counter party risk matter for the investors or not, see below for details. We write default risk is in the form $u = 1 - dk$ with 1 for not-defaulting and dk for the expected default rate.

The gain of the project can be written symbolically - i.e. without using summation and discounting notation, but focussing on the different parts in the gain function - as follow:

$$\text{Gain} = \begin{cases} \bar{I}, & \text{expected investment costs;} \\ dI, & \text{investment risk;} \\ \bar{c} \times \bar{p}, & \text{estimated savings (costs and volume);} \\ dc \times \bar{p}, & \text{volume risk;} \\ \bar{c} \times dp, & \text{energy price risk;} \\ dc \times dp, & \text{cross risk;} \\ dk \times c \times p, & \text{default risk.} \end{cases}$$

This defines the risk profile for the city without any structuring of risk. Therefore, the next question is: **Who bears which risk?** Professional technology provider keep the investment and volume risk due to their experience and their large project portfolio. That is variation in these two factors are absorbed in a large project portfolio. Consider an investor. The investor is willing to pay the expected investment costs $\bar{I}$ in exchange of participating at the future energy saving. That is, the city and the investor share future energy savings: The city participates with $\bar{c} \times \bar{p} \times a$ and the investor with $\bar{c} \times \bar{p} \times (1-a)$ at future energy savings. This defines the **performance contract**. The function a defines as function of time future participation. Since the investment has to be paid back to the investor, he will participate stronger at the beginning than the city. Else, the payback time increases. In this setup the whole investment is risk free for the city. The only risk which is not attributed is default risk of the city. Either it is passed and compensated to the investor or the bank keeps this risk. This type is a **structured product solution**. Other possible solutions are:

- City pays the project cash.
- City issues a bond.
- City issues a green bond.
- Bank issues a structured product (solution above).
- A special purpose vehicle is setup.

Before we consider some of these solutions we provide an example for the structured product. Assume a project which payback time 4y. Then the amount of saved energy $\bar{c} = 25\%$. Assume that the project costs 100 in a currency, that a increases linearly from 10 to 40 percent, $1 - a$ decreases linearly from 90 to 60 percent, that energy price risk is ± 2 percent per annum, that default risk of the city is 10 bps, that fees in structuring the deal are 1 percent per annum and that interest rates are flat at 2 percent.
Then,

- After 8y the whole energy savings belong to the city.
- After 5 years the investment amount is amortized, i.e. the years 6-8 generate return for the investor.
- The return for the investor is in case of constant energy prices equal to 6.3 percent, 5.3 percent if energy price fall by 2 percent each year and 7.1 percent in monotone increasing case. This return has to be corrected by the possible default of the city. If the investor does not want to take this default risk, the returns are lowered by the credit risk costs for the city.

Finally, if an investor wishes to get ride-off energy price risk the structuring delivers him fix energy prices or prices which are kept within a bandwidth.

From the other financing possibilities we only mention the green bond. This bond is issued by the city as an ordinary bond. The difference to such a bond is the coupon payment. The value of the coupon each year is determined by the price of the saved energy amount, i.e. it is a coupon derived from the underlying value 'energy price $\times$ saved energy volume'.

Clearly such a construction requires heavy legal and documentation work for and between the different parties. Furthermore, more hazard issues exists: The energy solution provider can change an excessive price I for the investment to cover possible price risk dI or the energy solution provider can predict biased low saved energy amounts to reduce its energy volume risk. To avoid such potential disincentives, a simple solution is let the energy firm itself invest into the project, i.e. to take a part of the investor's stake. This then both reduces moral hazard related to the investment amount and also to the expected energy volume savings since systematic deviations reduce the return of investment.

6.3 Demographic risk

6.3.1 Pension Economics

Pension funds face multiple requirements from different fields which are likely to be conflicting.

- Risk capacity. This capacity is determined for example by the age structure of the fund contributes, the actual asset and liability profile, the coverage ratio.
- Risk taking willingness. This follows from the investment goals, the expectations, the portfolio structure of the asset and liabilities.
- Risk-carry-duty. Depending on the jurisdiction and the type of the pension fund (defined benefit or defined contribution) the pension fund carries different risk such as a minimum guaranteed interest rate payment or a minimum return guarantee.
- Law and regulation. The depth and complexity of laws and regulations concerning pension funds differ for different jurisdictions. Pension funds are heavily regulated which can be understood as a precautionary measure to protect the savings of employees. But this hinders innovations in the pension scheme system and the whole retirement provision. Given the demographic change which many western countries face and the limited possibilities to intervene on a non-financial and contractual level, for example immigration, increasing the retirement age, increasing productivity, the need for financial innovations to adjust the risk- and return profiles in the retirement provision is evident.

A pension fund acting under these different forces should be (i) successful on a long term basis and (ii) the different parts of a fund - governing body, strategic committee, fund managers, controlling and audit units - should always be aware of the fact that the capital is entrusted to their care. The second point follows from the fact that although the fund itself does not faces a timing risk, the individual contributor is exposed to this risk, i.e. the date of retirement where labor income stops.

6.3.2 Capital Protected Investment

A main strategy to manage the fund capital is by direct investing into stocks, bonds or other basic financial instruments. The advantages of direct investments in stock markets are the full participation in rising markets and the flexibility. The disadvantages, which will lead to the risk transfer solution below, are that market volatility carries over to the pension fund assets. This affects earnings and risk budgeting, i.e. reactive behavior in the following sense is often observed. If markets boom, the earnings budget is increased which makes it more difficult to fulfill the budget in a sustainable way. If markets crash, the risk budget is often reduced which makes it more difficult to participate adequately if markets recover. Another drawback are investment decisions which depend heavily on market sentiments. First, media often exaggerate the sentiment, i.e. investment decisions are possibly exaggerated too. Second, diversity of opinion is lost, i.e. herding behavior is likely to follow. A final disadvantage are due to economic cycles. These cycles have a periodicity of several years and a rough rule of thumb is that on average two-third of the cycle is expansion and the rest contraction. On the other side, the board of directors of a pension fund expects a steady growing cover ratio or at least a constant one, low volatility of pension fund income and an appropriate interest yield. These expectations and the cyclically of the economic cycle lead to the conclusion that the investment earnings should be decoupled from the economic cycle. Figure 6.11 illustrates the discussion.

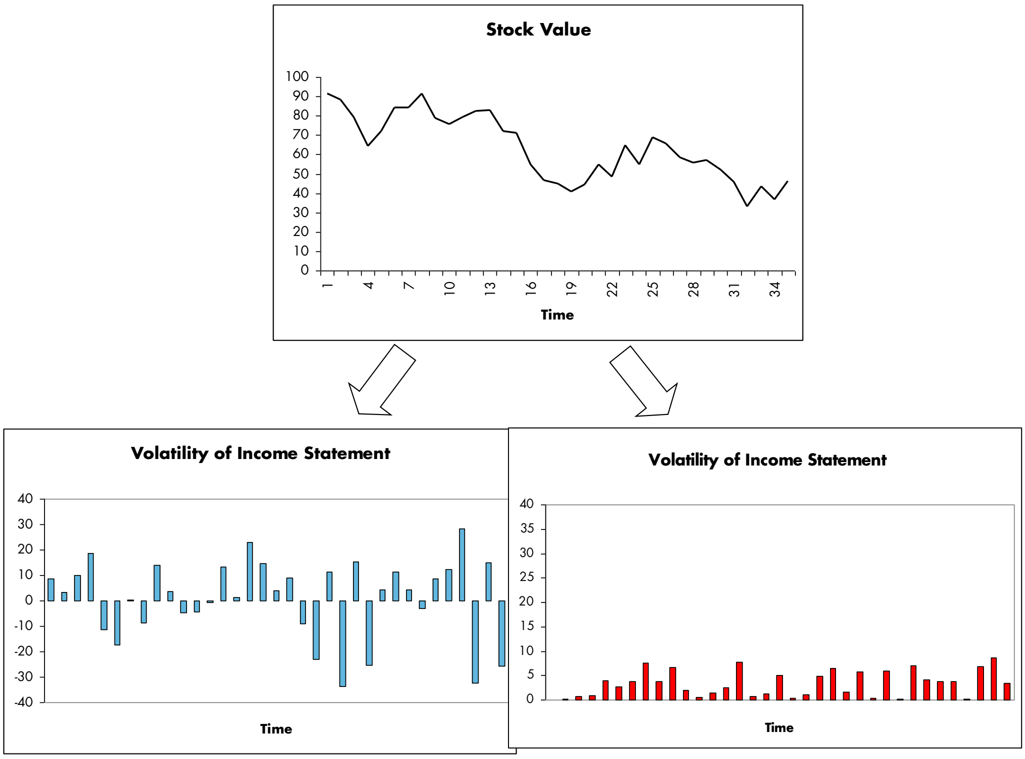


Figure 6.11: Income volatility due to a long-only position (Left Panel) and desired income volatility of a new solution (Right Panel).

In the left panel the income statement is volatile due to the direct investment in the risky assets without any risk transformation. The right panel shows a possible smoothing of income due to a risk transformation applied to the stock investment. The right panel cash flows lock like a floater. So why do we not simply consider a floater? The reason is that transforming volatility risk of stock markets into a floater-like cash flow stream the expected floating coupons are higher than using say LIBOR as underlying value.

How can such a positive cash flow stream obtained? Given the disadvantages of direct investment the solution is based on the following main issues:

- Rule based solution. This eliminates sentiment in decision making. The pension fund selects the underlying value.
- Long time to maturity to become independent of the economic life cycle, i.e. 10y or longer.
- Strategy
  - Full capital protection at maturity. It is difficult to argue given the importance of retirement capital to the contributors how avoiding capital protection can be justified.
  - Annual payment of coupons, this provides liquidity.
  - Risk transformation: Guarantee a minimum and a maximum coupon payment. Excess returns over the maximum coupon in specific years are kept in a reserve account - save in good years for possible bad years to come. Liquidate the reserve account at maturity which provides an additional coupon payment.

How can we achieve the desired risk transformation? Assume that we have $n = 0, 1, \ldots, T$ equidistant periods. At initiation, the reserve account $K_0 = 0$ is zero and the first coupon paid $c_1 = \bar{c}$ is equal to the maximum coupon $\bar{c}$, the cap level. Consider the second period. The driver which allows us to pay out coupons and to build reserves is the return $R_n = \frac{I_n - I_0}{nI_0}$ of the underlying index I. Note that the return is always calculated with respect initial index value. We compare the return at date $n = 2$, i.e. R_2 with the cap coupon. If the return is larger than the cap, $R_2 > \bar{c}$, we are in the comfortable situation to pay out the maximum coupon $c_2 = \bar{c}$ for the second time to the client and to take the excess to build up a reserve, i.e.

$$K_1 = K_0 + (R_1 - \bar{c})$$

where $K_0 = 0$. The above dynamics of the reserve account holds for all dates until maturity where $R_n > \bar{c}$, i.e.

$$K_n = K_{n-1} + R_n - \bar{c}.$$

If the return is smaller than the cap, $R_2 \leq \bar{c}$, we face a more interesting situation. Suppose that we pay out simply the return, i.e. $c_2 = R_2$ if R_2 is positive and zero else. Then we cannot build up a reserve in this period and we are still depending fully on the performance of the index in the next period. To get a more sensible rule we consider an arbitrary date n. At this date we face the following state variables, see Figure 6.12:

- The value of the reserve K_{n-1}.
- The return R_n.
- The cap level coupon $\bar{c}$.

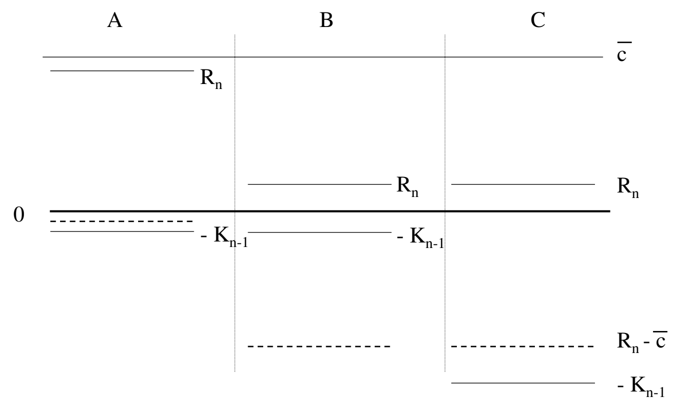


Figure 6.12: Case for the coupon in period n if the return of the index is lower than the cap value $\bar{c}$.

- R_n is not larger than $\bar{c}$.

The figure shows three cases A, B, C. In the case A the return R_n is close to the cap level but the reserve K_{n-1} are not very large. In B the return is still positive but close to zero. The reserve is still small. In C the difference is to B that we have collected a large reserve in last periods. Since $R_n \leq \bar{c}$ to fix the new coupon c_n we have to know how much we can take out of the potential reserve. Since $R_n - \bar{c} \leq 0$ we compare this figure with the reserve K_{n-1}. Suppose that $R_n - \bar{c} > -K_{n-1}$, i.e. either the return is close to the cap or there is large reserve account. Then we can take out $R_n - \bar{c}$ from the reserve. If this is not the case, we can only charge $-K_{n-1}$ to the reserve account. In summary, we charge at time n

$$B_n = \max(R_n - \bar{c}, -K_{n-1}) \leq 0$$

to the reserve account if the return is smaller than the cap. In this case the coupon is

$$c_n = \max(R_n + B_n, 0)$$

and the new reserve account value is

$$K_n = K_{n-1} + B_n \ .$$

The strategy so far contains a risk for the structuring firm that they fail to have enough capital at the end of the solution to cover its liability. Suppose that the index return is always lower than the cap coupon. Then $B_n = 0$ for all dates and the coupon payments are the full cap coupon after the first period and then the return is payed out in each period unless the return is negative. In this case, the derivative firm fails to finance the first, guaranteed cap coupon. Therefore, one introduces a condition that $c_n = \max(R_n + B_n, 0)$ is the new coupon in a period n only if the return and the charge of the reserve account are larger than a threshold value ac_{n-1} of the last coupon payment, i.e. if

$$R_n - B_n > ac_{n-1}$$

with $a < 1$. If this condition is not met, the payment is $c_n = ac_{n-1}$ and the reserve account is reset to zero, i.e. $K_n = 0$. Summarizing, the coupon payments and the new reserve account values at date n are given in next table:

	If $R_n > \bar{c}$	If $R_n \leq \bar{c}$& $R_n - B_n > ac_{n-1}$	If $R_n \leq \bar{c}$ & $R_n - B_n \leq ac_{n-1}$
Coupon	$c_n = \bar{c}$	$c_n = R_n + B_n$	$c_n = ac_{n-1}$
Reserve	$K_n = K_{n-1} + R_n - \bar{c}$	$K_n = K_{n-1} + B_n$	$K_n = 0$

Table 6.2: Coupons and reserve account formation.

In the final period, a fraction of the reserve account is payed out to the investor, i.e. he get the additional payoff

$$\tilde{c}_T = \min(K_T, \bar{c}) \ .$$

Example
Consider the data of Table 6.3

	0	1	2	3	4	5	6	7	8	9	10
Index I_t	100	85	95	130	148	145	142	142	136	145	98
R_t		-15	-2.5	10	12	9	7	6	4.5	5	-0.2
Min. Coupon		5	2.50	10	12	9	7	6	4.5	5	2.5
Max. Coupon			5	5	5	5	5	5	5	5	5
Reserve In-flow			0	5	7	4	2	1	0	0	0
Reserve Out-flow			0	0	0	0	0	0	-0.5	0	-5
Balance Reserve		0	0	5	12	16	18	19	18.5	18.5	13.5
Coupon paid		5	2.5	5	5	5	5	5	5	5	5

Table 6.3: Date for the example. Except for the first line all figures are in percentage.

The index first drops in the first period. It then recovers and even booms up to the second last period where it heavily drops from 145 to 98 in the last period. This paths defines the returns R_t. The maximum coupon is 5 percent. The first coupon paid is the maximum coupon - independent of realized returns. Since the index drops in the first and second period the reserve account cannot be raised, i.e. it is zero. The minimum coupon in period 2 of 2.5 percent is calculated as follows

$$c_n^{min} = \max(R_n, 0.5c_{n-1})$$

i.e. the contraction factor is $a = 0.5$. In period 3 a coupon of 4 percent is paid since $R_3 = 4\% > 0.5 \times 2.5\%$. At the end of period 4 for the first time the return is larger than the maximum coupon, i.e. we pay the maximum coupon an the excess is used to feed the reserve account. In period 5 we can pay the maximum coupon although the return is lower than the maximum coupon by taking 1 percent out of the reserve account. Although there is a heavy drop in the last period, the maximum coupon is paid thanks to the filled reserve account. This defines the logic and at the end of period 10 there is an additional coupon of 5 percent since the reserve is 5.5 percent and the minimum between the reserve and the maximum coupon is paid. The total value of the coupons is 46.5 percent.

The flow dynamics of the reserve account is as follows. The reserve inflow K_n^i at date n is given by

$$K_n^i = \max(R_n - c_n, 0) ,$$

the reserve outflow R_n^0 reads

$$K_n^o = \min(K_{n-1}, \min(c_n^{min}, c_n^{min} - \max(R_n, 0))) .$$

The balance reserve account K_n is given by

$$K_n = K_{n-1} + K_n^i + K_n^o .$$

The coupon paid at date n are given by

$$c_n = \max(c_n^{min}, R_n) - K_n^i - K_n^o . \tag{6.10}$$

c_n is a complicated recursion depending on R_n, c_{n-1}, a and K_{n-1}. The last dependence on the state variable K induces a path dependence, i.e. the coupon at date n depends on all former returns and all former paid coupons. The value of this cash flows is given by the expected value under a risk neutral probability. This cash flows defines option which belong to the cliquet family, i.e. option with a payoff

$$V_{\texttt{cliquet}}(T) = \min\left(C_{\texttt{glob}}, \max\left(F_{\texttt{glob}}, \sum_{j=1}^{M} \min\left(C_j, \max\left(b\left(\frac{I_j}{I_{j-1}} - K_j\right), F_j\right)\right)\right)\right) \tag{6.11}$$

with $F_{\texttt{glob}}$ the global floor, $C_{\texttt{glob}}$ the global cap, C_j and F_j the local cap and floor respectively and M the number of legs.

Figure 6.13 shows the coupon payments for the described structure of the product.

The underlying value is SMI and the period from 1960 up to 2000 is considered. The maximum coupon is equal to 5 percent. The structure is assumed to have a maturity of 10 years. It follows that the only period with significant low cash flows are the seventies of last century. In this phase stock markets showed very low volatility. There existed no source to generate large cash flows.

We compare the following pension fund investments:

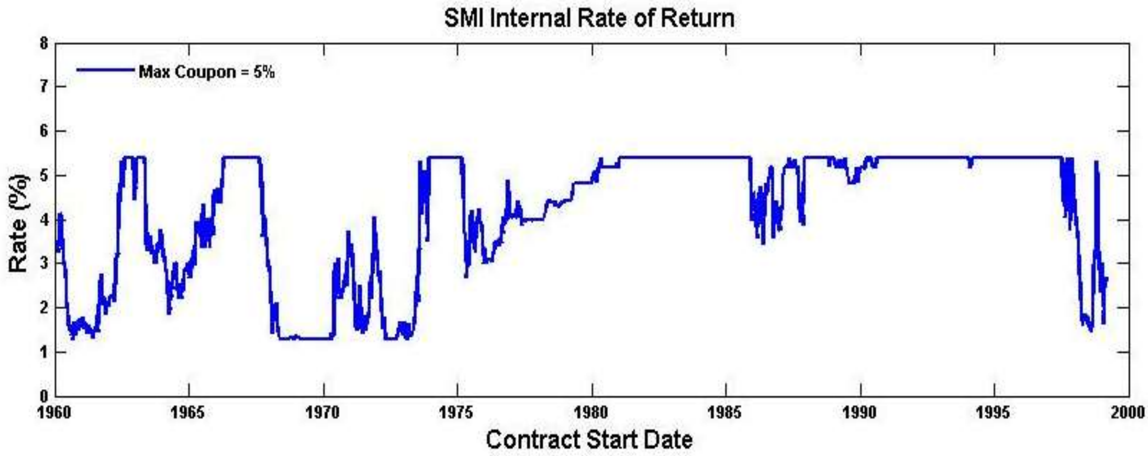


Figure 6.13: Backtesting of coupon payments for SMI. *Source: Zurich Cantonal Bank.*

- Direct investment in SMI versus
- a structured product which defines coupon payments as described above and where 100 percent capital protection is promised.

Consider Figure 6.14 in the sequel. The left panel shows the volatility of SMI during the last ten years as well the dividend yield. The right panel compares the two solutions. We us the cover ratio defined for the asset and liability value. We assume that the pension fund starts with a cover ratio of 100 as of December 1998. We assume that the stock quota is kept constant equal to 35 percent during the backtesting period and that dividends are reinvested. The other fraction of wealth is invested in other asset classes where fixed income is the dominant one. We assume that interest rate earning is equal to zero on all other asset classes than on SMI. This allows us to single out the difference between the two solution omitting interference with other factors.

The figure shows the stability of the cover ratio in the structured product solution, i.e. the effect of using a reserve account. This can also be seen as a strong decoupling of the economic life cycle. In the direct investment case volatility of the stock market

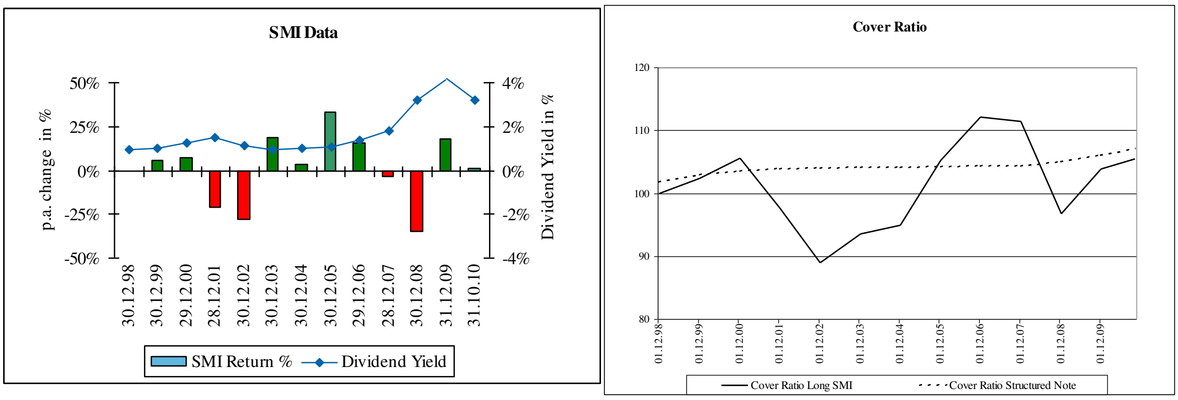


Figure 6.14: Historical SMI performance and dividend yield (Left Panel). Cover ratio for the direct investment and the structured product investment. *Source: Zurich Cantonal Bank.*

shows up directly in the cover ratio.

6.3.3 Risk Transfer: Intragenerational Cross Country Swaps and Longevity Risk

Countries at different stages of economic development face different long-term challenges. On the one hand, take a country such as Switzerland, which faces the challenge of being able to provide for the pensions and long-term care of its ageing population without jeopardizing its economic competitiveness. On the other hand, take Egypt, which faces the challenge of being able to attract sufficient funds and encourage new businesses to foster its economic development; Egypt is not constrained by longevity-related expenses, and higher longevity due to better health and living conditions may actually signify higher work productivity and faster economic growth. So Switzerland faces longevity risk, while Egypt faces growth risk. Rather than bear those risks, stakeholders in the two countries may benefit from transferring them to the financial markets. Yet the question is how.

Is there a possibility that rich and developing countries can exchange their risks such that a win-win situation follows? We show how such a swap functions which exchanges demographic risk and economic growth risk. This is a theoretical approach where we do not consider the many practical difficulties and risk in implementing such a swap.[a]

[a]This part is based on Padovani and Vanini (2010).

We delineate the structure of a potential future financial innovation enabling one party - the rich country - to mitigate its longevity risk and the other - the poor country - to widen its pool of funding options. This intergenerational cross-country swap would mirror the different timing of needed funds of the two countries and the funding capacities of each generation.

This swap permits risk management to be extended far beyond its former realm, covering a new class of risk: in this case, population ageing. It also changes the assumptions about what can be insured and hedged - take the old-age dependency ratio - and has a potential major impact on human welfare. A crucial aspect in this framework is the interrelationship between longevity and growth risks, since inadequate management of the high costs associated with an ageing society may lead a country to economic deterioration. The Standard & Poor's study documents how ageing-related government liabilities may result in downgrades of sovereign ratings if no adjustment in government budget occurs: Australia with AAA in 2005 is estimated to be non-investment grade in 2040. The same holds for France, US, UK.

The financial innovation we present is an agreement between a rich country's government and a poor country's government to exchange a sequence of cash flows at specified settlement dates. Its structure is designed in such a way to capture the temporal asymmetry between the two countries:

- the richer country needs funds in the future when it will have to cover expenses for the elderly,
- while the poorer country needs funds today to pay for educational, technological, and other infrastructure services.

The swap has a long-term duration, say 30 years as for available government debt. The rich country represents the fixed leg, the poor country the floating leg. The net cash flows from rich to poor are much higher in the first years of the contract's life, but this asymmetry is reversed over time, mirroring the different timing of needed and available funds stated above.

The swap's structure begs the question of how far ahead one can forecast growth and whether 30-year-ahead growth and longevity forecasts can be sufficiently reliable.

Given the actual difficulty in forecasting growth and longevity over a period exceeding 18 months, the swap is rolled over every 10 years. This also allows the parties to take into account sovereign default risk: If country creditworthiness declines, then the swap spread can be raised at the time the contract is rolled over. But as with any rolling strategy, it is important to account for possible rolling or tracking errors and their severity.

The main advantage for the rich country in this fixed-for-floating swap is to transform future cash-outs at a floating rate - its elderly-related expenses which depend on a stochastic population ageing rate - into payments at a fixed rate, thus locking in a 'sustainable' population ageing rate. As for the poor country, its main advantage is to hedge against adverse changes in the development aid it receives due to population ageing in the donor country. A long-term contract as this 30- year swap entails, though, the risk of significant fluctuations in exchange rates and interest rates. These risks may be hedged together by adding a (fixed-for-floating) cross-currency interest-rate swap, to be rolled over during the 30-year lifetime of the swap.

A couple of benefits of this intergenerational cross-country swap over a simple developmental loan from rich to poor are particularly relevant. One first benefit is the possibility for both parties to spread the cash out- and in flows over time. Secondly, the swap provides the possibility of including a rebate in case the contract is interrupted because of sovereign default or economic recession in either country. But this swap also offers advantages over a longevity derivative between two counter parties based in the rich country. The problem of finding a counter party willing to take on the long-longevity side of the transaction is often cited by practitioners as the main problem currently hindering the take-off of longevity derivatives. The issue of identifying the variable on which to base the swap's cash flows is not trivial, given that no existing measure to date serves our purpose fully and satisfactorily. We need a rate that reveals the strength with which longevity trends impact each country's wealth endowments. Gross domestic product (GDP) is by far the most widely-used indicator of economic growth, but it does have its shortcomings, too. A crucial shortcoming of GDP measures is that they do not take into account life expectancy or other demographic variables such as fertility, which are important indicators of a country's well-being. What is important for the purpose of this study, however, is to quantify the burden of elderly-related expenses on an economy. These expenses tend to actually raise GDP figures through an increase in government expenditures; but, as discussed above, longevity is forecast to negatively affect the economy of the more advanced economies. Several studies in growth theory attempt to come up with a growth measure reflecting the welfare gains from both quality and quantity of life. The aim of longevity-adjusted growth rates is to quantify the extent to which national economic growth is affected by population age structure. One major study in this field is that by Becker et al. (2006) who develop a statistical model that accounts for the impact of longevity on the evolution of welfare across almost 100 countries from 1960 to 2000. Their model measures the growth of individual income plus the value placed on the growth of an individuals life expectancy. A common result of all these studies is that, by not taking into account increases in longevity, GDP underestimates the extent to which developing countries are gaining relative to developed countries.

We follow a utility-based valuation approach, which relies on the individual rationality requirement of the agent's expected utility from participation in the contract to (weakly) exceed his reservation utility. We take the exponential utility function, which implies constant relative risk aversion and linear risk tolerance. This utility function is widely employed in the literature because of its mathematically tractability. In an incomplete market, to every contingent claim is associated an interval of arbitrage-free prices and arbitrage arguments alone are not sufficient to lead to a unique price, i.e. to a replication strategy. As the lower and upper endpoints of this interval coincide with the sub- and super-replication costs of the contingent claim, respectively, any price in the middle will lead to a possible profit & loss at maturity. Hence, the choice of an arbitrage-free price must be made with respect to another criterion. The pricing of the swap is based on the condition of individual rationality for both parties to enter the deal. The participation constraints allow us then to determine an interval of prices acceptable to both seller and buyer. The indifference valuation criterium demands that the investor valuing a contingent claim should achieve the same expected utility both in case he does not possess the claim and in case he does possess the claim but his initial capital is reduced by the amount of indifference value of the claim. In this framework, rich and poor country must be made indifferent between bearing longevity and income growth risks without entering a swap agreement under a medium population projection variant and no longer bearing these risks after entering the swap agreement.

If W_t denotes the wealth endowment of the rich government at time$t$ - national budget set aside for pensions and long term health care - adjusted for reference population size and in inflation and $\tilde{W}_t$ denotes the wealth endowment of the poor government at time t - national budget set aside for infrastructure projects - adjusted for reference population size and in inflation, u are the utility functions, D the discount factors, P^W the probability law of the wealth dynamics, P^a the probability law of the old-age dependency ratio[2] dynamics and $s_{bid/ask}$ respective fixed swap rates, the indifference values of the swap to the rich and the poor country are defined through (V is the value function)

$$V(W_0, s_{bid}, a_o) \geq V(W_0, 0, 0) \ , \ \tilde{V}(\tilde{W}_0, s_{ask}, a_o) \geq \tilde{V}(\tilde{W}_0, 0, 0) \ .$$

Using exponential utility, explicit bound for the swap rates follow.

[2]The old-age dependency ratio is the fraction of the population over 65 years old to the fraction of the population between 15 and 64 years old.

Chapter 7

Mathematical Appendix

7.1 Optimal Decision Making (Merton's Model)

The Principle of Optimality states: Let $0 < S < T$ with $[0, S]$ the initial condition. Then for all S selections, the control is optimal on $[S, T]$ with the starting value $W(S)$ which follows from the choice of c over $[0, S]$.

We first split the integral in two parts for small dt:

$$\begin{aligned} J(t_0, w_0) &= \max_c E\left[\int_{t_0}^{t_0+dt} u(t,c,W)dt + \int_{t_0+dt}^{T} u(t,c,W)dt + f(W(T),T)\right] \\ dW_t &= g(t,c,W)dt + \sigma(t,c,W)dB_t \ , \ W(t_0) = w_0 \ . \end{aligned} \tag{7.1}$$

Using the Principle of Optimality, the control function in the second integral should be optimal for the problem beginning at $t_0 + dt$ in the state $W(t_0 + dt) = w_0 + dW$. Hence,

$$\begin{aligned} J(t_0, w_0) &= \max_c E\left[\int_{t_0}^{t_0+dt} u(t,c,W)dt + \max_c E\left[\int_{t_0+dt}^{T} u(t,c,W)dt + f(W(T),T)\right]\right] \\ dW_t &= g(t,c,W)dt + \sigma(t,c,W)dB_t \ , \ W(t_0) = w_0 \ , \end{aligned} \tag{7.2}$$

or written with the value function (skipping the dynamics):

$$J(t_0, w_0) = \max_c E\left[\int_{t_0}^{t_0+dt} u(t,c,W)dt + J(t_0+dt, w_0+dW)\right] \ . \tag{7.3}$$

We next approximate the second value function since dt is small. This also allows us to assume that the control c is constant over a time interval with length dt. We get:

$$\begin{aligned} J(t_0, w_0) &= \max_c E[u(t,c,W)dt + J(t_0,w_0) + \partial_t J(t_0,w_0)dt \\ &+ \partial_w J(t_0,w_0)dW + 1/2\partial^2_{ww} J(t_0,w_0)(dW)^2] + o(dt) \ . \end{aligned} \tag{7.4}$$

This looks like a second order expansion in the state variable - but the square of Brownian motion $(dB)^2$) is linear in time (see the part and appendix on continuous time finance), i.e. $(dW)^2 = (g(t,u,W)dt + \sigma(t,u,W)dB)^2 = \sigma^2 dt$. The only random component in the above value function expression is therefore the term $\partial_w J dW$. Since $E[dB] = 0$, we get

$$E[\partial_w J dW] = \partial_w J g dt \ .$$

Dividing by dt we finally get the **fundamental partial differential equation (PDE)**

$$0 \quad = \quad \max_c \left[u + \partial_t J + \partial_w J g + +1/2 \partial^2_{ww} J \sigma^2 \right] . \tag{7.5}$$

Therefore,

- Taking formally the derivative w.r.t. to c in the above PDE gives us optimal decision making c as a function of the unknown value function J.
- Reinsert this candidate into the fundamental PDE (7.5) solve the resulting J-equation with the boundary and initial conditions (if any).
- Use this explicit solution J to obtain the fully specified optimal policy c_t^* and the optimal controlled state dynamics W_t^*.

We apply this to a specific model which was considered by Merton. The choice variable is a vector (c,ω) with the consumption rate and ω the fraction of wealth invested in the risky asset. The state variable W_t represents wealth. Utility index is

$$u(t,c,w) = e^{-rt} \frac{c^a}{a} \ , \ 0 < a < 1 \ .$$

We set the bequest motif $f(W(T),T) = 0$ equal to zero and assume that the individual optimizes his utility to infinity, i.e.

$$V(w_0) = \max_{c,\omega} E \left[\int_0^\infty e^{-rt} \frac{c_t^a}{a} dt \right] \ .$$

This infinite time problem leads to a slightly different fundamental PDE. To derive this, we set $J(t,W) = e^{-rt} V(W)$. Inserting this into the fundamental PDE leads after cancelling of the exponential function to

$$0 \quad = \quad \max_{c,\omega} \left[\frac{c^a}{a} - rV + \partial_w V g + +1/2 \partial^2_{ww} V \sigma^2 \right] . \tag{7.6}$$

The wealth dynamics W_t follows from the asset dynamics and the consumption rate. There is a risky asset with dynamics $dS/S = \mu dt + \sigma dB$ where the drift and the volatility are constant and a so-called risk less asset with dynamics $d\mathcal{B} = \mathcal{B} r dt$. The growth rate of wealth is the equal to the weighted sum of the asset growth rates minus the consumption rate, i.e.

$$dW/W = \omega dS/S + (1-\omega) d\mathcal{B}/\mathcal{B} - c/W dt \ .$$

The weight ω is equal to the number of risky assets times their price S divided by total wealth. Inserting the asset dynamics in the wealth growth rate equations gives the final wealth dynamics:

$$dW = (\omega\mu W + (1-\omega)rW - c)dt + \sigma\omega W dB \ .$$

Inserting this dynamics in the fundamental PDE gives:

$$0 \quad = \quad \max_{c,\omega}\left[\frac{c^a}{a} - rV + (\omega\mu W + (1-\omega)rW - c)\partial_w V + +1/2(\sigma\omega W)^2\partial^2_{ww}V\right] \ . \tag{7.7}$$

Taking the derivative w.r.t. to the two choice variables, setting them to zero gives the candidate solutions (First Order Conditions):

$$c^* = (\partial_w V)^{\frac{1}{a-1}} \ , \ \omega^* = \partial_w V \left(\frac{r-\mu}{\sigma^2}\frac{1}{W\partial^2_{ww}V}\right) \ . \tag{7.8}$$

This candidate optimal choice solution possess a drawback - they depend on the yet unknown value function. One has to determine the value function V. To achieve this, we reinsert the optimal candidate functions into the fundamental PDE. This gives an equation for the unknown value function V:

$$V = (\partial_w V)^{\frac{1}{a-1}}\frac{1-a}{a} + rW\partial_w V - \frac{(r-\mu)^2}{2\sigma^2}\frac{(\partial_w V)^2}{\partial^2_{ww}V} \ . \tag{7.9}$$

This is a highly non-linear equation and to find an analytical solution seems almost impossible. But we note that the value function $V(w)$ is proportional to the expected value of c^a. Therefore, a guess is to try $V(W) = \alpha W^a$ as a candidate solution with α a constant. Testing this guess in the PDE we see that all terms are proportional to W^a: We can factor out this power function times a complicated function which does not depend on the state variable W. Since this product has to be zero for all W, the complicated function has to be zero which gives us a value for the constant α and we obtained in this way a solution for the unknown value function. To carry this out we insert this guess into (7.9):

$$0 = W^a\alpha\underbrace{\left(\frac{1-a}{a}\alpha^{\frac{1}{a-1}} - 1 + ra - \frac{(r-\mu)^2}{2\sigma^2}\frac{a}{a-1}\right)}_{=:F(\alpha)} \ .$$

That is, the state variable dependence W^a appears in each term of the original PDE and can be factored out. Therefore $V(W) = \alpha W^a$ solves the PDE if $F(\alpha) = 0$. This equation can be solved explicitly, leading to a constant α^*. Hence we found a solution for the value function PDE which then provides us an explicit solution for the choice variables:

$$V(W) = \alpha^* W^a \quad , c^* = W(a\alpha^*)^{\frac{1}{a-1}} \ , \ \omega^* = \frac{\mu - r}{\sigma^2}\frac{1}{1-a} \ .$$

7.2 Volatility

We prove Dupire's equation (1.12).

Proof. The unknown density φ satisfies the Fokker-Planck equation:

$$\frac{1}{2}\frac{\partial^2}{\partial S_T^2}\left(\sigma^2 S_T^2 \varphi\right) - \frac{\partial}{\partial S_T}\left(r S_T \varphi\right) = \frac{\partial \varphi}{\partial T}.$$

Differentiating (1.11) w.r.t. K twice gives

$$\begin{aligned} \frac{\partial C}{\partial K} &= \frac{\partial}{\partial K}\int_K^\infty \varphi (S_T - K)\,\mathrm{d}S_T \\ &= \varphi(K-K) - \int_K^\infty \varphi\,\mathrm{d}S_T \\ &= -\int_K^\infty \varphi\,\mathrm{d}S_T \end{aligned}$$

and

$$\begin{aligned} \frac{\partial^2 C}{\partial K^2} &= -\frac{\partial}{\partial K}\int_K^\infty \varphi\,\mathrm{d}S_T \\ &= \varphi(K,T). \end{aligned}$$

Hence we can recover the risk neutral density φ from option data. Differentiating (1.11) w.r.t. T gives

$$\begin{aligned} \frac{\partial C}{\partial T} &= \int_K^\infty \left[\frac{\partial}{\partial T}\varphi(S_T,T)\right](S_T-K)\,\mathrm{d}S_T \\ &= \int_K^\infty \left[\frac{1}{2}\frac{\partial^2}{\partial S_T^2}\left(\sigma^2 S_T^2\varphi\right) - \frac{\partial}{\partial S}\left(r S_T \varphi\right)\right](S_T-K)\,\mathrm{d}S_T \end{aligned}$$

where we used the backward Fokker-Planck equation for the density function. Integrating by parts twice proves Dupires's equation. □

We prove the representation of local variance in terms of Black-Scholes implied variance, i.e. 1.14

Proof. Starting with the Black-Scholes implied volatility definition

$$C(S_0,K,T) = C_{\mathrm{BS}}(S_0,K,I(K,T),T),$$

and using the definitions x, y, the Black and Scholes formula becomes:

$$C(F_{0,T},x,y) = F_{0,T}\left(\Phi(d_1) - \mathrm{e}^x \Phi(d_2)\right),$$

with

$$d_1 = -\frac{x}{\sqrt{y}} + \frac{1}{2}\sqrt{y}, \quad d_2 = -\frac{x}{\sqrt{y}} - \frac{1}{2}\sqrt{y}.$$

The Dupire equation (1.12) becomes

$$\frac{\partial C}{\partial T} = \frac{\sigma^2_{\text{Loc}}}{2}\left[\frac{\partial^2 C}{\partial x^2} - \frac{\partial C}{\partial x}\right] + rC, \tag{7.10}$$

where $\sigma^2_{\text{Loc}} = \sigma^2_{\text{Loc}}(S_0, K, T)$ denotes the local variance. The derivatives of the Black-Scholes formula are

$$\frac{\partial^2 C_{\text{BS}}}{\partial y^2} = \left(-\frac{1}{8} - \frac{1}{2y} + \frac{x^2}{2y^2}\right)\frac{\partial C_{\text{BS}}}{\partial y},$$

$$\frac{\partial^2 C_{\text{BS}}}{\partial x \partial y} = \left(\frac{1}{2} - \frac{x}{y}\right)\frac{\partial C_{\text{BS}}}{\partial y},$$

and

$$\frac{\partial^2 C_{\text{BS}}}{\partial x^2} - \frac{\partial C_{\text{BS}}}{\partial x} = 2\frac{\partial C_{\text{BS}}}{\partial y}.$$

We may rewrite (7.10) in terms of implied variance by making the substitutions

$$\frac{\partial C}{\partial x} = \frac{\partial C_{\text{BS}}}{\partial x} + \frac{\partial C_{\text{BS}}}{\partial y}\frac{\partial y}{\partial x},$$

$$\frac{\partial^2 C}{\partial x^2} = \frac{\partial^2 C_{\text{BS}}}{\partial x^2} + 2\frac{\partial^2 C_{\text{BS}}}{\partial x \partial y}\frac{\partial y}{\partial x} + \frac{\partial^2 C_{\text{BS}}}{\partial y^2}\left(\frac{\partial y}{\partial x}\right)^2 + \frac{\partial C_{\text{BS}}}{\partial y}\frac{\partial^2 y}{\partial x^2},$$

and

$$\frac{\partial C}{\partial T} = \frac{\partial C_{\text{BS}}}{\partial T} + \frac{\partial C_{\text{BS}}}{\partial y}\frac{\partial y}{\partial T} = \frac{\partial C_{\text{BS}}}{\partial y}\frac{\partial y}{\partial T} + rC_{\text{BS}}.$$

Now (7.10) becomes

$$\begin{aligned}
\frac{\partial C_{\text{BS}}}{\partial y}\frac{\partial y}{\partial T} &= \frac{\sigma^2_{\text{Loc}}(K,T)}{2}\left[-\frac{\partial C_{\text{BS}}}{\partial x} + \frac{\partial^2 C_{\text{BS}}}{\partial x^2} - \frac{\partial C_{\text{BS}}}{\partial y}\frac{\partial y}{\partial x} + 2\frac{\partial^2 C_{\text{BS}}}{\partial x \partial y}\frac{\partial y}{\partial x}\right. \\
&\quad \left. + \frac{\partial^2 C_{\text{BS}}}{\partial y^2}\left(\frac{\partial y}{\partial x}\right)^2 + \frac{\partial C_{\text{BS}}}{\partial y}\frac{\partial^2 y}{\partial x^2}\right] \\
&= \frac{\sigma^2_{\text{Loc}}}{2}\frac{\partial C_{\text{BS}}}{\partial y}\left[2 - \frac{\partial y}{\partial x} + 2\left(\frac{1}{2} - \frac{x}{y}\right)\frac{\partial y}{\partial x}\right. \\
&\quad \left. + \left(-\frac{1}{8} - \frac{1}{2y} + \frac{x^2}{2y^2}\right)\left(\frac{\partial y}{\partial x}\right)^2 + \frac{\partial^2 y}{\partial x^2}\right].
\end{aligned}$$

Simplifying,

$$\frac{\partial y}{\partial T} = \sigma^2_{\text{Loc}}\left[1 - \frac{x}{y}\frac{\partial y}{\partial x} + \frac{1}{4}\left(-\frac{1}{4} - \frac{1}{y} + \frac{x^2}{y^2}\right)\left(\frac{\partial y}{\partial x}\right)^2 + \frac{1}{2}\frac{\partial^2 y}{\partial x^2}\right].$$

Finally, inverting this expresses the local variance as a function of the Black-Scholes implied variance $y = I^2_{BS}$, i.e.

$$\sigma^2_{\text{Loc}} = \frac{\frac{\partial y}{\partial T}}{1 - \frac{x}{y}\frac{\partial y}{\partial x} + \frac{1}{4}\left(-\frac{1}{4} - \frac{1}{y} + \frac{x^2}{y^2}\right)\left(\frac{\partial y}{\partial x}\right)^2 + \frac{1}{2}\frac{\partial^2 y}{\partial x^2}}.$$

Inserting the original variables and after some algebra the proof follows. $\square$

7.3 Riesz-Fischer Theorem, Separating Hyperplane Theorem, Fundamental Theorem of Finance

To prove the fundamental theorem of finance, we need the separating hyperplane theorem. We first state and prove the **Riesz-Fischer** theorem which states that any linear function on a vector space can be represented by an inner product.

7.3.0.1 Riesz-Fischer Theorem

Proposition 7.3.1 (Riesz-Fischer in $\mathbf{R}^n$). *Let $M \subset \mathbf{R}^n$ be a linear subset, $\langle \bullet, \bullet \rangle$ be the scalar product on $\mathbf{R}^n$, and $l : \mathbf{R}^n \to \mathbf{R}$ a linear map. Then there exists a unique vector $z \in \mathbf{R}^n$, such that $l(y)$ can be represented for all $y \in M$ by the scalar product $\langle z, y \rangle$, i.e.*

$$l(y) = \langle z, y \rangle \ , \ \forall z \in M. \tag{7.11}$$

Proof. We recall some facts from linear algebra first:
Let M and M' be subspaces of $\mathbf{R}^n$. Then M' **is the complement** *of M iff each vector $x \in \mathbf{R}^n$ can be written as the sum of two vectors $z \in M$ and $z' \in M'$ i.e.*

$$x = z + z' \ .$$

We write $\mathbf{R}^n = M \oplus M'$. Then M' is the complement of M iff M and M' have only the zero vector in common and the two spaces span $\mathbf{R}^n$, i.e.

$$\dim M + \dim M' = \dim \mathbf{R}^n = n \ .$$

If M is a linear subspace of $\mathbf{R}^n$, we define the orthogonal complement $M^\perp$:

$$M^\perp = \{x \in \mathbf{R}^n | \ \langle x, y, \rangle = 0 \ , \ \forall y \in M\} \ .$$

Then,

$$\mathbf{R}^n = M \oplus M^\perp \ .$$

If two vectors x, y are orthogonal, i.e. $\langle x, y, \rangle = 0$, the law of Phytagoras follows:

$$||x + y||^2 = ||x||^2 + ||y||^2 \ ,$$

where the norm $|| \bullet || = \sqrt{\langle \bullet, \bullet \rangle}$ is induced by the scalar product. Considering the orthogonal decomposition of $\mathbf{R}^n$ in $M \oplus M^\perp$, the vector $y \in M$ defined by

$$x = y + y' \ , y' \in M^\perp \ ,$$

is the orthogonal projection of x onto M. This vector has minimal distance to x, i.e.

$$y = \arg \min_{w \in M} ||x - w|| \ , \ \textit{if } x = y + y' \ , y' \in M^\perp, y \in M \ .$$

The last definitions from linear algebra are the kernel and the image of a linear map $f : \mathbf{R}^n \to \mathbf{R}^m$:

$$\ker f := \{x \in \mathbf{R}^n | \ f(x) = 0\} \subset \mathbf{R}^n$$

and

$$imf := \{y \in \mathbf{R}^m |\ y = f(x), x \in \mathbf{R}^n\} \subset \mathbf{R}^m \ .$$

The dimension formula

$$\dim \mathbf{R}^n = \dim \ker f + \dim imf$$

holds.

We start to prove the theorem:
If $l(y) = 0$, we set $z = 0$. Suppose that $l(y) \neq 0$. Since im$l \subset \mathbf{R}$, we have

$$\dim M = \dim \ker l + \dim \mathrm{im} l = \dim \ker l + 1 = \dim \ker l + \dim(\ker l)^{\perp} \ .$$

Since the kernel is a subspace it follows $\dim(\ker l)^{\perp} = 1$. Let $e \in M$ be a basis of $(\ker l)^{\perp}$. We decompose orthogonally the vector $y \in M$, i.e.

$$y = y' + \lambda e,\ y' \in \ker l, \lambda \in \mathbf{R} \ .$$

Since e and y' are orthogonal, we have:

$$\langle e, y\rangle = \langle e, y'\rangle + \lambda\langle e, e, \rangle = \lambda\langle e, e, \rangle$$

and therefore the multiple of λ equals

$$\lambda = \frac{\langle e, y\rangle}{\langle e, e, \rangle} \ .$$

For all $y \in M$ we get:

$$l(y) = l(y' + \lambda e) = l(y') + \lambda l(e) = \lambda l(e) \ ,$$

where we used the linearity of l and that $y' \in \ker l$. But this implies

$$l(y) = \lambda l(e) = \frac{\langle e, y\rangle}{\langle e, e, \rangle} l(e) = \langle \tilde{e}, y\rangle,\ \tilde{e} = \frac{l(e)e}{\langle e, e, \rangle} \ .$$

This proves, that each linear functional can be represented in the claimed form by a scalar product. The uniqueness is trivial by taking two different vectors $\tilde{e}$ and $\tilde{e}'$ which lead to the same representation of y and checking that the vectors have to be the same. $\square$

7.3.0.2 Separating Hyperplane Theorem

We first define the notion of a *hyperplane*. The intuition is the following: Let $U \subset \mathbf{R}^n$ a sub vector space, i.e. zero is in U. Translating this subspace produces a hyperplane H. Straight lines in $\mathbf{R}^2$ or planes in $\mathbf{R}^3$ which do not contain the zero vector are hyperplanes. They can be written in the form

$$\langle x, a\rangle = d \ .$$

Subspaces of $\mathbf{R}^n$ are kernels of linear maps l. The Riesz-Fischer theorem implies

$$l(x) = \langle x, a\rangle = 0$$

for the kernel. Since each affine subspace is representable by $l(x) = d$ we define:

Definition 7.3.2. *The hyperplane H_x through the vector x is defined by:*

$$H_x = \{a \in \mathbf{R}^n | \langle x, a \rangle = d\} \ , \tag{7.12}$$

and the half spaces $H_x^{+,-}$ are defined by:

$$H_x^{+,-} = \{a \in \mathbf{R}^n | \langle x, a \rangle \overset{\geq}{\leq} d\} \ . \tag{7.13}$$

Let U, V be two subsets of $\mathbf{R}^n$. The hyperplane H separates the sets $U, V \iff U$ and V are in different half spaces. The hyperplane H separates the sets U, V strictly $\iff$ H_x separates the sets and the sets are disjoint.

Proposition 7.3.3 (Separating hyperplane theorem)**.** *Let C and K be two disjoint and convex subsets of $\mathbf{R}^n$. Let C be compact and K be closed. Then there exits a hyperplane H, which separates C and K strictly.*

The compactness of one set is necessary:

$$U = \{(x, y) \in \mathbf{R}^2 |\ x > 0, y \geq \frac{1}{x}\} \ , \ V = \{(x, y) \in \mathbf{R}^2 |\ x > 0, y \geq -\frac{1}{x}\} \ .$$

The sets are disjoint and convex. But they are not compact and therefore they cannot be strictly separated (draw a figure).

We show that there exists a hyperplane through z_0, which is perpendicular to y_0x_0 and which does not intersect U, V. Let

$$d(C, K) = \inf_{x \in C, y \in K} ||x - y||$$

be the shortest distance between C and K. It is proven in analysis courses that for C compact and K closed such minimizing points x_0, y_0 exist, i.e.

$$d(C, K) = ||x_0 - y_0|| > 0 \ .$$

Let H_{x_0} be the hyperplane through x_0, which is perpendicular to y_0x_0. We write H_{x_0} as follows:

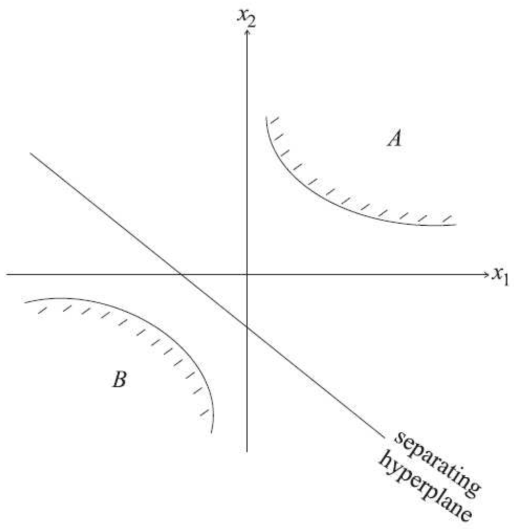
$$H_{x_0} = \{z \in \mathbf{R}^n | \langle y_0 - x_0, z - x_0 \rangle = 0\} \ .$$

Let $\phi(\lambda)$ be the function, which measures the distance between y_0 and x:

$$\begin{aligned} \phi(\lambda) &:= ||y_0 - (x_0 + \lambda(x - x_0))||^2 \\ &= \langle y_0 - x_0, y_0 - x_0 \rangle - 2\lambda \langle y_0 - x_0, x - x_0 \rangle + \lambda^2 \langle x - x_0, x - x_0 \rangle. \end{aligned}$$

This function is continuously differentiable and we have $\phi(\lambda) \geq \phi(0)$, $\forall \lambda \in [0, 1]$, since x_0 is closest to y_0. Therefore, $\phi'(\lambda) = -2\langle y_0 - x_0, x - x_0 \rangle + 2\lambda \langle x - x_0, x - x_0 \rangle$ and

$$\phi'(0) = -2\langle y_0 - x_0, x - x_0 \rangle \geq 0 \ ,$$



Proof.

Figure 7.1: The hyperplane separates the two convex sets A and B in $\mathbf{R}^2$. A set is convex if any 'line with end and starting point in the set remains fully in the set'.

i.e.

$$\langle x_0 - y_0, x - x_0 \rangle \leq 0 \ , \ \forall x \in U,$$

since C is convex. In the same way one shows that for H_{y_0} the inequality

$$\langle y_0 - x_0, y - y_0 \rangle \leq 0 \ , \ \forall x \in U$$

holds. Since, for all $y \in V$

$$\langle y_0 - x_0, y - y_0 \rangle = \langle y_0 - x_0, y_0 - x_0 \rangle + \langle y_0 - x_0, y - y_0 \rangle \geq 0$$

holds. It follows that H_{x_0} separates the sets C, V and the same is true for H_{y_0}. Therefore, H_{z_0} separates the sets strictly. □

7.3.1 Proof of the First Fundamental Theorem of Finance

We prove the Proposition 2.3.8

Proof. $\Rightarrow$. The first step is to prove the following claim: Let ψ be a vector where all components are strictly positive. This is a state vector, if each attainable claim or payoff $V = \mathbb{P}\phi$ implies $\langle \psi, V \rangle = \langle S_0, \phi \rangle$ (we omit the time index T). $V = \mathbb{P}\phi$ implies with the use of transposition rule for matrices

$$\langle \psi, V \rangle = \langle \psi, \mathbb{P}\phi \rangle = \langle \mathbb{P}'\psi, \phi \rangle.$$

If ψ is a state vector we have $S_0 = \mathbb{P}'\psi$, i.e. $\langle \psi, V \rangle = \langle S_0, \phi \rangle$ follows. If $\langle \psi, V \rangle = \langle S_0, \phi \rangle$, we get $\langle \mathbb{P}'\psi, \phi \rangle = \langle S_0, \phi \rangle$, i.e. $S_0 = \mathbb{P}'\psi$. This proves the claim.
Using this, we know that for each attainable payoff $V = \mathbb{P}\phi$ the identity $\langle \psi, V \rangle = \langle S_0, \phi \rangle$ holds. Therefore, if all components of V are positive, also $\langle S_0, \phi \rangle \geq 0$ holds. If all components of the payoff are strictly positive, the same applies to $\langle S_0, \phi \rangle$. Therefore, arbitrage is not possible.

$\Leftarrow$. This part of the proof needs the Separating Hyperplane Theorem: For K, M be to closed, disjoint convex sets in $\mathbb{R}^d$ where K is also compact sets there exists a non-zero vector z and a real number b such that for all $x \in M, y \in K$ the inequality

$$\langle z, x \rangle < b < \langle z, y \rangle$$

holds. We apply this result and set

$$M = \{(x, x_{K+1}) \in \mathbb{R}^{K+1} | \ x = \mathbb{P}\phi, x_{K+1} = -\langle S_0, \phi \rangle = -V_0\}$$

and

$$K = \{x \in \mathbb{R}^{K+1} | \ x_i \geq 0, \sum_i x_i = 1\} \ .$$

M is an augmented space of payoffs. It consists of all payoffs at date T plus the price of the portfolio $-\langle S_0, \phi \rangle$ at time zero. K is a simplex (the boundary of a pyramid in three dimensional space). Both sets are closed and K is compact. M is a subspace of the corresponding Euclidian space, i.e. convex and closed. Since the compact set lies in the positive orthant the definition of no arbitrage implies that M and K are disjoint. The Separation Theorem then applies:

There exists a vector $z \in \mathbb{R}^{K+1}$ such that $\langle z, x\rangle < b < \langle z, y\rangle$ for all $x \in M, y \in K$.

Since M is a linear space, this inequalities can only hold if the vector z is element of the orthogonal complement $M^{\perp}$. But then the number b is strictly positive. Since also $\langle z, y\rangle > b > 0$ for $y \in K$, all components of the vector z are strictly positive. This allows us to define the state price density as: $\psi_k = \frac{z_k}{z_{K+1}}$ for all k. We show that this is indeed a state price vector, i.e. each component is strictly positive (which follows from the property of the vector z) and that ψ solves $S_0 = \mathbb{P}'\psi$. To prove this recall that z is element of the orthogonal complement of M and therefore for each strategy vector $\phi \in \mathbb{R}^N$ we have:

$$0 = \langle z_{K+1}\psi, \mathbb{P}\phi\rangle - z_{K+1}\langle S_0, \phi\rangle = z_{K+1}(\langle \mathbb{P}'\psi, \phi\rangle - \langle S_0, \phi\rangle) \ .$$

Therefore, $\langle \mathbb{P}'\psi, \phi\rangle = \langle S_0, \phi\rangle$, i.e. $\mathbb{P}'\psi = S_0$, holds for all strategies ϕ. This proves the claim. $\square$

7.4 Proofs of the Cox-Ross-Rubinstein (CRR) Model

We prove Proposition 2.4.4

Proof. We first show that the X_k's are i.i.d. under Q. We know

$$E^Q[X_{k+1}|\mathcal{F}_k] = 1 + r.$$

which is equivalent to

$$\begin{aligned} E^Q[X_{k+1}|\mathcal{F}_k] &= E^Q[X_{k+1}\chi_{\{X_{k+1}=1+u\}}|\mathcal{F}_k] + E^Q[X_{k+1}\chi_{\{X_{k+1}=1+d\}}|\mathcal{F}_k] \\ &= (1+u)E^Q[\chi_{\{X_{k+1}=1+u\}}|\mathcal{F}_k] + (1+d)E^Q[\chi_{\{X_{k+1}=1+d\}}|\mathcal{F}_k] = 1 + r \end{aligned}$$

with the indicator function $\chi_{\{X_{k+1}=1+u\}} = 1$ if $X_{k+1} = 1 + u$ and else χ is zero. Since X_k can assume either $1 + u$ or $1 + d$, we have

$$E^Q[\chi_{\{X_{k+1}=1+u\}}|\mathcal{F}_k] + E^Q[\chi_{\{X_{k+1}=1+d\}}|\mathcal{F}_k] = 1.$$

We set $q_k := E^Q[\chi_{\{X_{k+1}=1+u\}}|\mathcal{F}_k]$ and $1-q_k := E^Q[\chi_{\{X_{k+1}=1+d\}}|\mathcal{F}_k]$. Then $q_k+1-q_k = 1$ follows and also

$$uq_k + d(1 - q_k) = r \ .$$

Since this equations hold true for all k, q_k does not depend upon k. Hence, we set $q := q_k$ and $1 - q := 1 - q_k$. Solving this equations

$$q = \frac{r-d}{u-d} \; , \; 1 - q = \frac{u-r}{u-d} \, .$$

Therefore,

$$q_k = q_0 = q = E^Q[\chi_{\{X_{k+1}=1+u\}}] = Q[X_{k+1} = 1 + u]$$

and

$$\begin{aligned} Q[X_{k+1} &= 1 + u|\mathcal{F}_k] = Q[X_{k+1} = 1 + u] = q \\ Q[X_{k+1} &= 1 + d|\mathcal{F}_k] = Q[X_{k+1} = 1 + d] = 1 - q. \end{aligned}$$

This shows that the X_k's are identically distributed. To show independence, we calculate the law for $x_1, x_2 \in \{1 + u, 1 + v\}$:

$$\begin{aligned} \hat{\pi}[X_1 = x_1, X_2 = x_2] &= \hat{\pi}[X_1 = x_1]Q[X_2 = x_2|X_1 = x_1] \\ &= \hat{\pi}[X_1 = x_1]Q[X_2 = x_2] \; . \end{aligned}$$

This is inductively generalized to

$$\hat{\pi}[X_1 = x_1, X_2 = x_2, \ldots, X_T = x_T] = \hat{\pi}[X_1 = x_1]Q[X_2 = x_2] \cdots Q[X_T = x_T],$$

which implies independence of X_{k+1} under Q.

Assuming conversely, that the random variables X_k are i.i.d. with

$$Q[X_k = 1 + u] = q = \frac{r-d}{u-d} \; , \; Q[X_k = 1 + d] = 1 - q = \frac{u-r}{u-d} \; ,$$

it follows $E[X_k] = 1 + r$ and

$$E^Q[X_{k+1}|\mathcal{F}_k] = E^Q[X_{k+1}] = 1 + r \implies E^Q[S_{k+1}|\mathcal{F}_k] = (1 + r)S_k \; .$$

This proves that $\tilde{S}_k$ is a Q-martingale, since

$$E^Q[\tilde{S}_{k+1}|\mathcal{F}_k] = \tilde{S}_k \; .$$

□

Proof of Proposition 2.4.5

Proof. We use the following trick for $t < T$

$$S_T = S_t \frac{S_{t+1}}{S_t} \cdots \frac{S_T}{S_{T-1}} = S_t \prod_{i=t+1}^{T} X_i \; .$$

The reason to do this is the conditional expectation we have to calculate for pricing the call option, i.e. since the random variables X_k are i.i.d. under Q, this will considerably simplify the calculations. We get

$$\tilde{C}_t = (1+r)^{-T} E^Q[(S_T - K)_+ | \mathcal{F}_t] = (1+r)^{-T} E^Q[(S_t \prod_{i=t+1}^{T} X_i - K)_+ | \mathcal{F}_t] \ .$$

Consider the random variables $\prod_{i=t+1}^{T} X_i$ are known, i.e.

$$\prod_{i=t+1}^{T} X_i = (1+u)^m (1+d)^{(T-t)-m}$$

where m is the number of upwards move in $(t,T]$. The probability that exactly $m (\leq (T-t))$ upwards moves are realized is under Q independent of $\mathcal{F}_t$ and equals the binomial distribution

$$Q[\prod_{i=t+1}^{T} X_i = (1+u)^m (1+d)^{(T-t)-m}] = \binom{T-t}{m} \hat{p}^m \hat{q}^{(T-t)-m} \ ,$$

where we used the i.i.d. property[1]. We get

$$\begin{aligned} C_t^* &= (1+r)^{-T} E^Q[(S_T - K)_+ | \mathcal{F}_t] = (1+r)^{-T} E^Q[(S_t \prod_{i=t+1}^{T} X_i - K)_+ | \mathcal{F}_t] \\ &= (1+r)^{-T} \sum_{m=0}^{T-t} (S_t (1+u)^m (1+d)^{(T-t)-m} - K)_+ \binom{T-t}{m} \hat{p}^m \hat{q}^{(T-t)-m}. \end{aligned}$$

Hence, C_t can be written as a function of S_t and t: $C_t = c(S_t, t)$ with

$$c(x,t) = (1+r)^{-(T-t)} \sum_{m=0}^{T-t} (x(1+u)^m (1+d)^{(T-t)-m} - K)_+ \binom{T-t}{m} \hat{p}^m \hat{q}^{(T-t)-m}.$$

The term $(x(1+u)^m(1+d)^{(T-t)-m} - K)_+$ increases with m, and is positive, i.e.

$$\left(\frac{1+u}{1+d}\right)^m > \frac{K}{x(1+d)^{(T-t)}} \ ,$$

has to hold. Setting

$$m_0 := \frac{\ln(\frac{K}{x}) - (T-t)\ln(1+d)}{\ln(1+u) - \ln(1+d)},$$

[1] The best way to think about $\binom{T-t}{m}$ is the following one: How many different possibilities are there to put m indistinguishable balls into $T-t$ boxes? Well, $\binom{T-t}{m}$ possibilities. In the CRR model, the balls are the upwards move in $(t,T]$ when it order of the moves is irrelevant.

and denoting with $\lceil m_0 \rceil$ the next larger natural number of m_0, we get

$$\begin{aligned}
c(x,t) &= (1+r)^{-(T-t)} \sum_{m=\lceil m_0 \rceil}^{T-t} (x(1+u)^m(1+d)^{(T-t)-m} - K) \binom{T-t}{m} \hat{p}^m \hat{q}^{(T-t)-m} \\
&= x(1+r)^{-(T-t)} \sum_{m=\lceil m_0 \rceil}^{T-t} \binom{T-t}{m} (\hat{p}(1+u))^m (\hat{q}(1+d))^{(T-t)-m} \\
&\quad -K(1+r)^{-(T-t)} \sum_{m=\lceil m_0 \rceil}^{T-t} \binom{T-t}{m} \hat{p}^m \hat{q}^{(T-t)-m} \\
&= x \sum_{m=\lceil m_0 \rceil}^{T-t} \binom{T-t}{m} \left(\frac{\hat{p}(1+u)}{1+r}\right)^m \left(\frac{\hat{q}(1+d)}{1+r}\right)^{(T-t)-m} \\
&\quad -K(1+r)^{-(T-t)} \sum_{m=\lceil m_0 \rceil}^{T-t} \binom{T-t}{m} \hat{p}^m \hat{q}^{(T-t)-m} \\
&= x\bar{B}(T, p^*, \lceil m_0 \rceil) - K(1+r)^{-(T-t)} \bar{B}(T, \hat{p}, \lceil m_0 \rceil)
\end{aligned}$$

where $p^* := \frac{\hat{p}(1+u)}{1+r}$ and $\hat{p} = \frac{r-d}{u-d}$. Denoting with $B(T,p,m)$ the cumulative distribution function of a binomial distributed random variable with parameters T and p evaluated at m, we then have $\bar{B}(T,p,m) := 1 - B(T,p,m-1)$.

It follows that the price at time zero of a call is ($C_0 = c(S_0, 0)$):

$$C_0 = S_0 \bar{B}(T, p^*, \lceil m_0 \rceil) - K(1+r)^{-T} \bar{B}(T, \hat{p}, \lceil m_0 \rceil).$$

This proves the Binomial Option-Pricing Formula.

□

We prove Proposition 2.4.7.

Proof. We calculate the expectation and the variance of $\ln(\frac{S_T^*}{S_0})$ under Q.

In the m-period model

$$\ln(\frac{\tilde{S}_T}{S_0}) = \ln(\prod_{i=1}^{m} \tilde{X}_i) = \sum_{i=1}^{m} \ln \tilde{X}_i \ ,$$

where

$$\tilde{X}_i = \frac{\tilde{S}_i}{\tilde{S}_{i-1}} \in \left\{ \frac{1+u_m}{1+r_m}, \frac{1+d_m}{1+r_m} \right\} \qquad (i = 1, \ldots, m)$$

and $Q[\tilde{X}_i = \frac{1+u_m}{1+r_m}] = q_m$. The X_i's are under Q i.i.d. and $(\tilde{S}_k)_{k=0,\ldots,m}$ are a Q-martingale. For $\ln(\tilde{X}_i)$ follows

$$\ln(\tilde{X}_i) \in \left\{ +\sigma\sqrt{\frac{T}{m}}, -\sigma\sqrt{\frac{T}{m}} \right\} \ ,$$

$$E^Q[\ln(\tilde{X}_i)] = (2q_m - 1)\sigma\sqrt{\frac{T}{m}}$$

and

$$E^Q[(\ln(\tilde{X}_i))^2] = \sigma^2\frac{T}{m} \ .$$

We next calculate the expected value $E^Q[\ln(\tilde{X}_i)]$. Writing

$$(1+u_m) \quad = \quad (1+r_m)e^{\sigma\sqrt{\frac{T}{m}}} \ , \ (1+d_m) = (1+r_m)e^{-\sigma\sqrt{\frac{T}{m}}} \ ,$$

it follows

$$\begin{aligned}
u_m - d_m &= (1+r_m)(e^{\sigma\sqrt{\frac{T}{m}}} - e^{-\sigma\sqrt{\frac{T}{m}}}) \\
r_m - d_m &= (1+r_m)(1 - e^{-\sigma\sqrt{\frac{T}{m}}}) \\
q_m &= \frac{r_m - d_m}{u_m - d_m} = \frac{1 - e^{-\sigma\sqrt{\frac{T}{m}}}}{e^{\sigma\sqrt{\frac{T}{m}}} - e^{-\sigma\sqrt{\frac{T}{m}}}} \ .
\end{aligned}$$

This implies

$$\begin{aligned}
E^Q[\ln(\tilde{X}_i)] &= (2q_m - 1)\sigma\sqrt{\frac{T}{m}} \\
&= \frac{2 - 2e^{-\sigma\sqrt{\frac{T}{m}}} - e^{\sigma\sqrt{\frac{T}{m}}} + e^{-\sigma\sqrt{\frac{T}{m}}}}{e^{\sigma\sqrt{\frac{T}{m}}} - e^{-\sigma\sqrt{\frac{T}{m}}}}\sigma\sqrt{\frac{T}{m}} \\
&= \frac{2 - e^{-\sigma\sqrt{\frac{T}{m}}} - e^{\sigma\sqrt{\frac{T}{m}}}}{e^{\sigma\sqrt{\frac{T}{m}}} - e^{-\sigma\sqrt{\frac{T}{m}}}}\sigma\sqrt{\frac{T}{m}} \ .
\end{aligned}$$

Developing the fraction in a Taylor series, we get

$$(2q_m - 1) = \frac{-\sigma^2\frac{T}{m}}{2\sigma\sqrt{\frac{T}{m}}} + o(1/m) = -\frac{\sigma}{2}\sqrt{\frac{T}{m}} + o(1/m) \ .$$

The expression $o(1/m)$ is the *Landau o-Symbol's* from analysis. Therefore,

$$E^Q[\ln(\tilde{X}_i)] = -\frac{\sigma^2}{2}\frac{T}{m} + o(1/m^{3/2})$$

and

$$Var(\ln(\tilde{X}_i)) = \frac{\sigma^2 T}{m} - \frac{\sigma^4 T}{2m^2} + o(1/m^2) \ .$$

Independence implies

$$
\begin{aligned}
E^Q[\ln \frac{\tilde{S}_T}{S_0}] &= \sum_{i=1}^{m} E^Q[\ln(\tilde{X}_i)] \\
&= mE^Q[\ln(\tilde{X}_1)] = -\frac{\sigma^2 T}{2} + o(1/\sqrt{m}) \\
Var(\ln \frac{\tilde{S}_T}{S_0}) &= \sum_{i=1}^{m} Var(\ln(\tilde{X}_i)) = mVar(\ln(\tilde{X}_1)) \\
&= \sigma^2 T + o(1/\sqrt{m}).
\end{aligned}
$$

For m large, the expected value of $\ln(S_T^*/S_0)$ approaches the constant $-\frac{\sigma^2 T}{2}$ and the variance of $\ln(\tilde{S}_T/S_0)$ tends towards $\sigma^2 T$. The proof uses characteristic functions (Fourier transform). The characteristic function $\Psi_{X(t)} := E[e^{iXt}]$ uniquely determines the probability distribution of the random variable X (i.e. there is a one-to-one relation between probability distributions and characteristic functions) and if the sequence $\Psi_{X_n}(t)$ converges point wise in t to $\Psi_X(t)$, then X_n converges in probability to X. We calculate the characteristic function of $\ln(\tilde{S}_T/S_0)$:

$$
\begin{aligned}
\Psi_{\ln(\tilde{S}_T/S_0)}(t) &= E[e^{it\sum_{k=1}^{m}\ln(\tilde{X}_k)}] = E[\prod_{k=1}^{m} e^{it\ln(\tilde{X}_k)}] = \prod_{k=1}^{m} E[e^{it\ln(\tilde{X}_k)}] \\
&= \Big(E[e^{it\ln(\tilde{X}_1)}]\Big)^m = \left(E[1 + it\ln(X_1^*) - \frac{t^2\ln(\tilde{X}_1)^2}{2} + o(1/m)]\right)^m \\
&= \left(1 + itE[\ln(\tilde{X}_1)] - \frac{t^2 E[(\ln(\tilde{X}_1)^2]}{2} + o(1/m)\right)^m \\
&= \left(1 + \frac{it\frac{-\sigma^2 T}{2} - \frac{t^2\sigma^2 T}{2}}{m} + o(1/m)\right)^m \\
&\longrightarrow e^{it\frac{-\sigma^2 T}{2} - \frac{t^2}{2}\sigma^2 T} \quad (m \to \infty).
\end{aligned}
$$

But this is the characteristic function of a normally distributed random variable with mean $-\frac{\sigma^2 T}{2}$ and variance $\sigma^2 T$ where we used twice independence of $\ln(\tilde{X}_k)$.

□

We proof proposition 2.4.8.

Proof. We calculate the price of a put option in the m-model first. We know that

$$
P_0^{(m)} = E^Q[\Big(K(1+r_m)^{-m} - \tilde{S}_T\Big)_+] = E^Q[\Big(K(1+r_m)^{-m} - S_0 e^{\ln(\tilde{S}_T/S_0)}\Big)_+] .
$$

To take the limit $m \to \infty$ in the put option formula means to interchange the integral limit (expected value) with the $m \to \infty$ limit. Since the function under the expectation is bounded, we can interchange the limits, i.e.[2]

$$\lim_{m\to\infty} P_0^{(m)} = E[(Ke^{-RT} - S_0 e^Y)_+] ,$$

where Y is a normally distributed random variable with mean $-\sigma^2 T/2$ and variance $\sigma^2 T$. Setting $X = \frac{Y+\sigma^2 T/2}{\sigma\sqrt{T}}$ implies that X is standard normally distributed with density function $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. This implies

$$\begin{aligned}
\lim P_0^{(m)} &= E[(Ke^{-RT} - S_0 e^{\sigma\sqrt{T}X} e^{-\frac{\sigma^2 T}{2}})_+] \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (Ke^{-RT} - S_0 e^{\sigma\sqrt{T}x} e^{-\frac{\sigma^2 T}{2}})_+ e^{-\frac{x^2}{2}}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} (Ke^{-RT} - S_0 e^{\sigma\sqrt{T}x} e^{-\frac{\sigma^2 T}{2}}) e^{-\frac{x^2}{2}}\, dx \\
&= Ke^{-RT} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{x^2}{2}}\, dx - S_0 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{\sigma\sqrt{T}x} e^{-\frac{\sigma^2 T}{2}} e^{-\frac{x^2}{2}}\, dx \\
&= Ke^{-RT}\Phi(a) - S_0 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}(x^2 - 2\sigma\sqrt{T}x + \sigma^2 T)}\, dx \\
&= Ke^{-RT}\Phi(a) - S_0 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{(x-\sigma\sqrt{T})^2}{2}}\, dx \\
&= Ke^{-RT}\Phi(a) - S_0 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a-\sigma\sqrt{T}} e^{-\frac{u^2}{2}}\, du \\
&= Ke^{-RT}\Phi(a) - S_0 \Phi(a - \sigma\sqrt{T}) .
\end{aligned}$$

The constant a is determined by the equation:

$$Ke^{-Rt} = S_0 e^{\sigma\sqrt{T}a} e^{-\frac{\sigma^2 T}{2}}$$

which implies $a = \frac{\ln(K/S_0) - RT}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{t}}{2}$. Φ denotes the distribution functions of a standard normally distributed random variable , i.e.

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}}\, dx .$$

The put-call parity implies $C_0^{(m)} = P_0^{(m)} + S_0 - K(1 + r_m)^{-m}$ and we get for the fair price

$$\begin{aligned}
\lim_{m\to\infty} C_0^{(m)} &= S_0(1 - \Phi(a - \sigma\sqrt{T})) - Ke^{-RT}(1 - \Phi(a)) \\
&= S_0\Phi(-a + \sigma\sqrt{t}) - Ke^{-RT}\Phi(-a)
\end{aligned}$$

Denoting $d_1 := -a + \sigma\sqrt{t}$ and $d_2 := -a$ we proved the claim.

□

[2]The conditions of the Lebesgue dominated convergence theorem are satisfied.

7.5 Brownian Motion

We define and prove the **quadratic variation** property of Brownian motion.

Proposition 7.5.1. *The quadratic variation of Brownian motion satisfies:*

$$[W, W](t) := \lim_{n\to\infty} \pi_n(W) = t$$

in the square-integrability (L^2) sense, where

1. $\pi_n(W) = \sum_{t_i\in\pi_n} \left(W_{t_{i+1}} - W_{t_i}\right)^2$ *with* $(\pi_n)_n$ *a refining sequence of partitions of* $[a, a+t]$*, i.e.* $\pi_m \subset \pi_n$ *for* $m > n$*, and*

$$\lim_{n\to\infty} |\pi_n| = \lim_{n\to\infty} \max_i (t^n_{i+1} - t^n_i) = 0 \ .$$

Proof. We get from the definition

$$\pi_n(W) - t = \sum_{t_i\in\pi_n} \left(W_{t_{i+1}} - W_{t_i}\right)^2 - (t_{i+1} - t_i) = \sum_{t_i\in\pi_n} Y_i$$

with (Y_i) a sequence if independent random variables with mean zero. Taking expectations we have

$$E[(\pi_n(W) - t)^2] = \sum_{t_i\in\pi_n} E[Y_i^2] \ .$$

Since $\frac{W_{t_{i+1}} - W_{t_i}}{t_{i+1} - t_i}$ has the distribution of $Z^2 \sim \mathcal{N}(0, 1)$ we obtain:

$$\begin{aligned} E[(\pi_n(W) - t)^2] &= \sum_{t_i\in\pi_n} E[Y_i^2] = E[(Z^2 - 1)^2] \sum_{t_i\in\pi_n} (t_{i+1} - t_i)^2 \\ &\leq E[(Z^2 - 1)^2]|\pi_n| \to 0 \ (n \to \infty) \end{aligned}$$

since the partition norm tends to zero. This proves L^2-convergence. □

Consider as an example a continuous deterministic function f instead of the Brownian motion. The quadratic variation $[f, f] = 0$ follows, since $(f_{ti+1} - f_{t_i})^2$ vanishes if $|\pi_n| \to 0$. The first variation of a continuous function is also always zero. For a stochastic process the first variation is defined as follows.

Definition 7.5.2. *The first variation* $V_A(\omega)$ *of a stochastic process* X *on a interval* $A = [a, b]$ *for a path* ω *is defined by*

$$V_A(\omega) = \sup_{\pi\in\mathcal{P}} \sum_{t_i\in\pi} |X_{t_{i+1}}(\omega) - X_{t_i}(\omega)| \tag{7.14}$$

with $\mathcal{P}$ *the set of all finite partitions of* $[a, b]$.

The next property shows a crucial property of the Brownian motion.

Proposition 7.5.3. *For almost all $\omega \in \Omega$, the path $t \to W_t(\omega)$ of the Brownian motion are of unbounded variation.*

Proof. Suppose $P(V_A < \infty) > 0$. Let (π_n) a refining sequence with norm tending zero. From the proof of the last theorem we have on the set $\{\omega | \ V_A(\omega) < 0\}$ where the variation is finite:

$$\begin{aligned} b - a &= \lim_{n\to\infty} \sum_{t_i \in \pi_n} (W_{t_{i+1}} - W_{t_i})^2 \\ &= \lim_{n\to\infty} \sup_{t_i \in \pi_n} |W_{t_{i+1}} - W_{t_i}| \sum_{t_i \in \pi_n} |W_{t_{i+1}} - W_{t_i}| \\ &= \lim_{n\to\infty} \sup_{t_i \in \pi_n} |W_{t_{i+1}} - W_{t_i}| V_A \to 0 \ (n \to \infty) \end{aligned}$$

since $\sup_{t_i \in \pi_n} |W_{t_{i+1}} - W_{t_i}| \to 0$ for $|\pi_n| \to 0$ by continuity of the paths of the Brownian motion. Hence, $b - a < 0$ follows which is absurd. □

7.6 Geometric Brownian Motion, Ornstein-Uhlenbeck, Integration by Parts, Black and Scholes

We first solve the geometric Brownian motion dynamics, i.e. we prove (2.83). For $X_t \neq 0$, we have

$$\int_0^t \frac{dX_t}{X_t} = rt + \sigma W_t \ . \tag{7.15}$$

The problem is that we do not know how to integrate $\int_0^t \frac{dX_t}{X_t}$. If the equation would be ordinary differential equation, than the logarithm would be a solution. We use this as a guess for Itô's formula. That is, we set $f(t,x) = \ln x$. Using the multiplication rule we get

$$\begin{aligned} d(\ln X) &= \frac{1}{X} dX_t + \frac{1}{2}(-\frac{1}{X^2})(dX)^2 \\ &= \frac{1}{X} dX - \frac{1}{2}\frac{1}{X^2}(rXdt + \sigma X dW)^2 = \frac{1}{X}dX - \frac{1}{2}\frac{1}{X^2}\sigma^2 X^2 dt. \end{aligned}$$

This implies $\frac{dX}{X} = d(\ln X) + \frac{1}{2}\sigma^2 dt$. Integrating,

$$\int_0^t \frac{dX_t}{X_t} = x + rt + \sigma W_t = \ln X_t + \frac{1}{2}\sigma^2 t.$$

Finally the solution of (**??**) reads

$$X(t) = X_0 e^{(r - \frac{1}{2}\sigma^2)t + \sigma W_t} \ . \tag{7.16}$$

To check that this is indeed the solution, apply Itô's formula and verify that (**??**) is satisfied.

We next calculate the expected value of the solution X_t, i.e. what is

$$E_P[X_t] = X_0 e^{(\mu - \frac{1}{2}\sigma^2)t} E[e^{\sigma W_t(\omega)}]? \tag{7.17}$$

To calculate the integral, we use Itô's formula for the function $Y_t := e^{\sigma W_t}$. We get

$$dY_t = \sigma e^{\sigma W_t} dW_t + \frac{1}{2}\sigma^2 e^{\sigma W_s} dt$$

and therefore

$$Y_t = Y_0 + \sigma \int_0^t e^{\sigma W_s} dW_s + \frac{1}{2}\sigma^2 \int_0^t e^{\sigma W_s} ds.$$

Taking expectations of this last equation implies

$$E[Y_t] = E[Y_0] + \sigma E[\int_0^t e^{\sigma W_s} dW_s] + \frac{1}{2}\sigma^2 E[\int_0^t e^{\sigma W_s} ds],$$

where the middle term is zero. We get

$$E[Y_t] = Y_0 + \frac{1}{2}\sigma^2 \int_0^t E[Y_s] ds,$$

which is an ordinary differential equation

$$\frac{d}{dt} E[Y_t] = \frac{1}{2}\sigma^2 E[Y_s]$$

with the solution

$$E[Y_t] = e^{\frac{1}{2}\sigma^2 t}.$$

Hence, we obtain for the expected value of the geometric Brownian motion

$$E[X_t] = X_0 e^{\mu t}. \tag{7.18}$$

This prove **??**. The proof of the variance is similar.

We solve the *Ornstein-Uhlenbeck* SDE

$$dX_t = -cX_t dt + \sigma dW_t \ , \ X_0 = x, \tag{7.19}$$

with σ, c constants. To find a solution we use the partial integration formula. We guess that $Y_t = X_t e^{ct}$ leads to a solution since this would be a guess for the deterministic equation without a Brownian motion part. We use the partial integration formula to determine the differential dY:

$$\begin{aligned}
dY_t &= dX_t e^{ct} + cX_t e^{ct}dt + \underbrace{ce^{ct}dX_t dt}_{=0} \\
&= dX_t e^{ct} + cX_t e^{ct}dt = (-cX_t dt + \sigma dW_t)e^{ct} + cX_t e^{ct}dt \\
&= e^{ct}\sigma dW_t \\
&\Longrightarrow \\
Y_t &= X_t e^{ct} = Y_0 + \int_0^t e^{cs}\sigma dW_s \\
&\Longrightarrow \\
X_t &= e^{-ct}Y_0 + e^{-ct}\int_0^t e^{cs}\sigma dW_s \ .
\end{aligned}$$

Since $Y_0 = X_0 = x$ we finally have

$$X_t = e^{-ct}x + e^{-ct}\int_0^t e^{cs}\sigma dW_s \ . \tag{7.20}$$

Since the expected value of the Itô Integral w.r.t. to the Brownian motion is zero we get

$$E[X_t] = e^{-ct}x \to 0 \ (t \to \infty) \tag{7.21}$$

and we get for the variance:

$$\begin{aligned}
var(X_t) &= E[(X_t - E[X_t])^2] = E[(e^{-ct}x + e^{-ct}\int_0^t e^{cs}\sigma dW_s - e^{-ct}x)^2] \\
&= E[(e^{-ct}\int_0^t e^{cs}\sigma dW_s)^2] \\
&\quad \text{using the Itô isometry we get} \\
&= E[e^{-2ct}\int_0^t (e^{cs}\sigma)^2 ds] = \sigma^2 e^{-2ct}E[\int_0^t e^{2cs}ds] \\
&= \sigma^2 e^{-2ct}(e^{2ct} - 1)\frac{1}{2c} = \frac{\sigma}{2c}(1 - e^{-2ct}) \ .
\end{aligned}$$

We prove the **integration by parts formula** (2.69):

Proof. We set: $f(X,Y) = (Y+X)^2, f_1(X) = X^2$ and $f_2(Y) = Y^2$ and we apply Itô's formula to this functions:

$$\begin{aligned}
df &= \partial_X f dX + \partial_Y f dY + \frac{1}{2}(\partial^2_{XX}(dX)^2 + \partial^2_{YY}(dY)^2 + 2\partial^2_{XY}dXdY) \\
&= 2(X-Y)dX + 2(Y+X)dY + \frac{1}{2}2dX^2 + \frac{1}{2}2dY^2 + 2dXdX \\
&= 2(X+Y)(dX+dY) + \sigma^2 dt + \sigma'^2 dt + 2\sigma\sigma' dt \\
&= 2(X+Y)(dX+dY) + (\sigma+\sigma')^2 dt \ .
\end{aligned}$$

Since $df_1(X) = dX^2 = 2Xdx + \sigma^2 dt$ and similar for f_2 we get

$$df - df_1 - df_2 = 2d(XY) = 2XdY + 2YdX + 2\sigma\sigma' dt$$

i.e. the claim is proven. □

We check that:

$$d\tilde{V}_t^{\xi} = \psi_t d\tilde{S}_t \; . \tag{7.22}$$

Proof.

$$\begin{aligned} d\tilde{V} &= d(\frac{S}{B}) = \frac{1}{B}dV - \frac{V}{B^2}dB \\ &= \frac{1}{B}(\phi dB + \psi dS_1) - \frac{\phi B}{B^2}dB - \frac{\psi S}{B^2}dB \\ &= \frac{\psi dS}{B} - \frac{\psi S}{B^2}dB = \psi d\tilde{S} \; . \end{aligned}$$

□

We prove the **call option price** formula in Black and Scholes model.

Proof. Consider $X = f(S_t) = (S_t - K)^+$. Starting from the representation of the payoff, we get

$$\begin{aligned} V_t &= E_Q[e^{-r(T-t)} f(S_T)|\mathcal{F}_t] \\ &= E_Q[e^{-r(T-t)} f(Se^{r(T-t)} e^{\sigma(\tilde{W}_T - \tilde{W}_t) - \frac{1}{2}\sigma^2(T-t)})|\mathcal{F}_t]. \end{aligned}$$

Since S_t is known for $\mathcal{F}_t$ and the increment $\tilde{W}_T - \tilde{W}_t$ is under Q a centered Gaussian with variance $T - t$ we get

$$V_t = e^{-r(T-t)} \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} f(Se^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma y\sqrt{T-t}}) e^{-\frac{1}{2}y^2} dy \; .$$

Using the definition of the call and defining $d_2 = d_1 - \sigma\sqrt{\tau}, \quad d_1 = \frac{\log\frac{S}{K} + (r+\frac{1}{2}\sigma^2)\tau}{\sqrt{\tau}}$ we get

$$\begin{aligned} V_t &= \frac{1}{\sqrt{2\pi}} \int_{d_2}^{\infty} (xe^{-\frac{1}{2}\sigma^2\tau + \sigma\sqrt{\tau}y} - e^{-r\tau}K)e^{-\frac{1}{2}y^2} dy \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d_2} (xe^{-\frac{1}{2}\sigma^2\tau - \sigma\sqrt{\tau}y} - e^{-r\tau}K)e^{-\frac{1}{2}y^2} dy \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d_2} xe^{-\frac{1}{2}z^2} dz - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d_2} e^{-r\tau}K)e^{-\frac{1}{2}y^2} dy, \quad (z = y + \sigma\sqrt{\tau}) \\ &= x\ \Phi(d_1) - Ke^{-r(T-t)}\ \Phi(d_2). \end{aligned}$$

But this is the same expression we derived for a call option considering the limit of the Cox-Ross-Rubinstein model. □

7.7 Markowitz

Proof of positive definite property 3.23:

Proof. Consider a model with only two risky assets and let $\phi = (\phi_1, 1-\phi_1)$ be a normalized portfolio. Suppose further that for V

$$\langle \phi, V\phi \rangle < 0 \tag{7.23}$$

for some portfolio ϕ. This is equivalent to

$$\phi_1^2 \mathrm{var}(R_1) + (1-\phi_1)^2 \mathrm{var}(R_2) + 2\phi_1(1-\phi_1)\mathrm{cov}(R_1, R_2) < 0 \ .$$

Rearranging this last expression in powers of the portfolio strategy, the following equivalent condition is obtained

$$\begin{aligned} 0 \quad > \quad & \phi_1^2 \underbrace{(\mathrm{var}(R_1) + \mathrm{var}(R_2) - 2\mathrm{cov}(R_1, R_2))}_{=:x} \\ + \quad & 2\phi_1 \underbrace{(\mathrm{cov}(R_1, R_2) - \mathrm{var}(R_2))}_{=:y} + \underbrace{\mathrm{var}(R_2)}_{=:z} \ . \end{aligned}$$

The RHS of this inequality is a parabola in ϕ_1. Since $x > 0$ the parabola obtains the minimum at $\phi_1^* = -\frac{y}{x}$. Inserting this last result in the RHS of the above inequality the equivalent inequality is obtained:

$$-\frac{y^2}{x} + z < 0 \ .$$

Since $x > 0$, this condition is equivalent to

$$y^2 > zx \ .$$

Reinserting the definition of x, y, z, we get

$$\frac{\mathrm{cov}(R_1, R_2)}{\mathrm{var}(R_1)\mathrm{var}(R_2)} > 1 \ ,$$

i.e. a contradiction to the properties of the correlation coefficient. Therefore, condition (7.23) cannot be true and the covariance matrix V needs to satisfy for all portfolio strategies ϕ the inequality

$$\langle \phi, V\phi \rangle \geq 0 \ , \ \forall \phi \ ,$$

i.e. V has to be *positive semi definite*. □

Proof of Proposition 3.1.3:

Proof. To prove the statement we have for an equally-distributed portfolio strategy $\phi_i = \frac{1}{N}$:

$$\begin{aligned}\langle\phi, V\phi\rangle &= \sum_{i=1}^{N}\phi_i^2 var(R_i) + \sum_{i,j=1,i\neq j}^{N}\phi_i\phi_j cov(R_i,R_j)\\ &= (\frac{1}{N})^2 N\frac{1}{N}\sum_{i=1}^{N}\mathrm{var}(R_i)\\ &\quad +(\frac{1}{N})^2(N^2-N)\frac{1}{N^2-N}\sum_{i,j=1,i\neq j}^{N} cov(R_i,R_j)\\ &= \frac{1}{N}\overline{\sigma}_N^2 + (\frac{1}{N})^2(N^2-N)\overline{\mathrm{cov}}_N\\ &= \frac{1}{N}\overline{\sigma}_N^2 + (1-\frac{1}{N})\overline{\mathrm{cov}}_N,\end{aligned}$$

where we introduced the mean variance $\overline{\sigma}_N^2$ and the mean covariance $\overline{\mathrm{cov}}_N$. □

Proof of Proposition 3.7.1:

Proof. With the Lagrangian

$$L = \frac{1}{2}\langle\phi, V\phi\rangle + \lambda_1(1-\langle e,\phi\rangle) + \lambda_2(r-\langle\mu,\phi\rangle),$$

the first order conditions

$$0 = \frac{\partial L}{\partial\phi} := \begin{pmatrix}\frac{\partial L}{\partial\phi_1}\\ \frac{\partial L}{\partial\phi_2}\\ \vdots\\ \frac{\partial L}{\partial\phi_N}\end{pmatrix},$$

are a set of N equations. From

$$\frac{1}{2}\frac{\partial\langle\phi, V\phi\rangle}{\partial\phi} = V\phi\ ,\quad \frac{\partial\langle\phi,\mu\rangle}{\partial\phi} = \mu$$

we get

$$0 = V\phi - \lambda_1 e - \lambda_2\mu \tag{7.24}$$
$$1 = \langle e,\phi\rangle \tag{7.25}$$
$$r = \langle\mu,\phi\rangle\ . \tag{7.26}$$

The optimality conditions are therefore $N+2$ linear equations in the $N+2$ variables $\phi, \lambda_1, \lambda_2$. To solve this system, we proceed as follows. Since V is strictly positive definite, V^{-1} exists and (7.25) implies

$$\phi = \lambda_1 V^{-1}e + \lambda_2 V^{-1}\mu\ .$$

Multiplying this last equation from the left with e and μ, respectively, and using the normalization condition and the return constraint, we get a linear system for the two Lagrange multipliers:

$$\begin{aligned} 1 &= \lambda_1 \langle e, V^{-1} e\rangle + \lambda_2 \langle e, V^{-1}\mu\rangle \\ r &= \lambda_1 \langle \mu, V^{-1} e\rangle + \lambda_2 \langle \mu, V^{-1}\mu\rangle \ . \end{aligned}$$

If we set $\tau = (\lambda_1, \lambda_2)'$ and $y = (1, r)'$ the last system reads

$$y = \begin{pmatrix} \langle e, V^{-1} e\rangle & \langle e, V^{-1}\mu\rangle \\ \langle \mu, V^{-1} e\rangle & \langle \mu, V^{-1}\mu\rangle \end{pmatrix} \tau =: A\tau \ . \tag{7.27}$$

If A is invertible, we are done since then $y = A\tau$ can be trivially solved for τ. This determines the Lagrange multipliers λ_i^* and inserting this result in $\phi^* = \lambda_1^* V^{-1} e + \lambda_2^* V^{-1}\mu$ gives us the optimal portfolio and proves the proposition. We prove that within the given model, the matrix A is invertible, i.e. we claim that $\det A = \Delta > 0$. To prove this we use the Cauchy-Schwartz inequality: For two arbitrary vectors x, y we have

$$|x|^2 |y|^2 \geq \langle x, y\rangle^2 \ ,$$

where the strict inequality holds if the two vectors are independent. To rewrite the determinant in the form needed for the Cauchy-Schwartz inequality, we first define the vectors x, y. That for, we use the decomposition $V = WW'$, which always exists for strictly positive definite, symmetric matrices. Using this, we get

$$\langle e, V^{-1} e\rangle = \langle e, (WW')^{-1} e\rangle = \langle e, (W')^{-1} W^{-1} e\rangle = \langle W^{-1} e, W^{-1} e\rangle =: \langle x, x\rangle \ ,$$

where we used

$$\langle x, A' A x\rangle = \langle Ax, Ax\rangle$$

and properties of the matrix inverse. Proceeding in the same form with the other elements of A and defining $\langle \mu, V^{-1} e\rangle = \langle y, x\rangle$ we get

$$\det A = \langle x, x\rangle \langle y, y\rangle - \langle x, y\rangle^2 \ .$$

Using the Cauchy-Schwartz inequality $\det A \geq 0$ follows. Since μ and e are linearly independent, the same holds for $x = W^{-1} e$ and $y = W^{-1}\mu$ too. This proves $\det A > 0$ and the prove of the proposition is completed. For reference, we note the optimal multiplier values:

$$\begin{aligned} \lambda_1^* &= (A^{-1} y)_1 = \frac{1}{\Delta}\left(-\langle \mu, V^{-1}\mu\rangle + r\langle e, V^{-1}\mu\rangle\right) \qquad (7.28) \\ \lambda_2^* &= (A^{-1} y)_2 = \frac{1}{\Delta}\left(-\langle e, V^{-1}\mu\rangle + r\langle e, V^{-1} e\rangle\right) \ . \end{aligned}$$

□

Proof of Proposition 3.7.2:

Proof. Since V is strictly positive definite, $\langle e, V^{-1}e\rangle > 0$ follows. We next use the spectral theorem of linear algebra:

Proposition 7.7.1. *If V is a positive definite, symmetric matrix, there exists and orthogonal matrix*[3] *U such that $U'VU = \Lambda$, with Λ a diagonal matrix and the eigenvalues of V as its entries.*

Using this proposition, we set $e = Uy$ with U a $N \times N$ orthogonal matrix and we get

$$\begin{aligned} b = \langle e, V^{-1}e\rangle &= \langle y, U'V^{-1}Uy\rangle = \sum_i \lambda_i^{(V^{-1})} y_i^2 \\ &\leq \lambda_{max}^{(V^{-1})}\langle y, y\rangle = \lambda_{max}^{(V^{-1})}\langle e, U'Ue\rangle = \lambda_{max}^{(V^{-1})}|e|^2 \\ &= \frac{1}{\lambda_{min}} N, \end{aligned}$$

where $U'V^{-1}U$ is a diagonal matrix and where for the last step we used the following calculation. Consider the two-dimensional case, i.e.

$$U'V^{-1}U = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} .$$

Using the definition of orthogonal matrices,

$$U'V^{-1}U = U^{-1}V^{-1}U'^{-1} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}^{-1} = \frac{1}{\lambda_1\lambda_2}\begin{pmatrix} \lambda_2 & 0 \\ 0 & \lambda_1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\lambda_1} & 0 \\ 0 & \frac{1}{\lambda_2} . \end{pmatrix}$$

Similar, a lower bound follows for the smallest eigenvalue which proves

$$\langle e, V^{-1}e\rangle \in [\lambda_{min}^{(V^{-1})}|e|^2, \lambda_{max}^{(V^{-1})}|e|^2] = [\frac{N}{\lambda_{max}}, \frac{N}{\lambda_{min}}] .$$

The proof of the second claim is analogous. To prove the last inequality, we first note that the matrix

$$A = (e, \mu)'V(e, \mu) = \begin{pmatrix} \langle e, V^{-1}e\rangle & \langle e, V^{-1}\mu\rangle \\ \langle e, V^{-1}\mu\rangle & \langle \mu, V^{-1}\mu\rangle \end{pmatrix}$$

is invertible due to the independence of the vectors e and μ. The Cauchy-Schwarz inequality implies that $\det A > 0$ which is equivalent to

$$|\langle e, V^{-1}\mu\rangle| \leq \sqrt{\langle \mu, V^{-1}\mu\rangle\langle e, V^{-1}e\rangle} \leq \frac{\sqrt{N}|\mu|}{\lambda_{min}} .$$

□

Proof of the Mutual Fund Theorem (Proposition 3.7.3).

[3] A real matrix U is orthogonal, if and only if $U' = U^{-1}$ and $\det U = \pm 1$.

Proof. The KKT (Karush-Kuhn-Tucker= conditions in the proof of Proposition 3.7.1) imply that any solution of the optimization problem is of the form

$$\phi = \lambda_1 V^{-1} e + \lambda_2 V^{-1} \mu \ .$$

The first term is proportional to the global minimum variance portfolio; the second one is proportional to $\hat{\phi}^*$. This proves the first claim. To prove the second claim, assume that ϕ^1 and ϕ^2 are two minimum variance portfolios. Then, they can be spanned by the global minimum variance portfolio and the portfolio $\hat{\phi}^*$ due to the first part of the proposition as follows

$$\phi^i = (1 - a^i)\phi^*_m + a^i \hat{\phi}^* \ , \ i = 1, 2 \ .$$

A solution ϕ^* of the minimum-variance problem can then be written as

$$\phi^* = \frac{\lambda_1 b + a^2 - 1}{a^2 - a^1} \phi^1 + \frac{1 - a^1 - \lambda_1 b}{a^2 - a^1} \phi^2 \tag{7.29}$$

where the multiplier λ_1 is given in (7.28). The coefficients of the above representation add up to 1. □

Proof of Proposition .

Proof. To prove this, note that the vector of the covariances of the returns with the portfolio is obtained by calculating $V\phi$. To prove necessity, we assume that the portfolio ϕ is efficient. It follows from (3.26

$$V\phi = \frac{rb - c}{\Delta}\mu + \frac{a - rc}{\Delta} e$$

i.e.

$$\mathrm{cov}(R_i, R^\phi) \quad = \quad \frac{rb - c}{\Delta}\mu_i + \frac{a - rc}{\Delta} = \frac{bE[R^\phi] - c}{\Delta} E[R_i] + \frac{a - cE[R^\phi]}{\Delta} \ . \tag{7.30}$$

We verify $\frac{bE[R^\phi]-c}{\Delta} \geq 0$. Linear independence of the vectors μ, e and the Cauchy-Schwarz inequality imply $\Delta > 0$. Since $E[R^\phi] \geq \frac{c}{b}$, necessity is proven. For the minimum variance portfolio we note $E[R^{\phi^*_{min}}] = \frac{c}{b}$. To prove sufficiency, we assume that (3.32) holds true. In vector notation this reads

$$V\phi = f_1^\phi \mu + f_2^\phi e \ , \ f_i^\phi \geq 0 \ . \tag{7.31}$$

The weights f_i^ϕ of this portfolio are obtained by multiplication from the left with V^{-1} and using the restriction $\langle \mu, \phi \rangle = E[R^\phi]$ and the normalization condition. Solving the two equation with respect to f_1^ϕ, f_2^ϕ and inserting the results into (3.32) implies that (3.26) is satisfied with $E[R^\phi] \geq \frac{c}{b}$. □

The **Markowitz optimization problem with a risk less asset** reads:

$$\begin{array}{lll} \min\limits_{\phi} & \frac{1}{2}\langle \phi, V\phi\rangle & \mathcal{M}_R \\ \text{s.t.} & \langle e, \phi\rangle = 1-\phi_0 & \\ & \langle \mu, \phi\rangle = r-\mu_0\phi_0\ . & \end{array} \tag{7.32}$$

The solution of this problem together with some properties are summarized in the next proposition.

Proposition 7.7.2. *If the same assumptions as in Proposition 3.7.1 hold, then the model $\mathcal{M}_R$ possesses the solution*

$$\begin{aligned} \phi^* &= \lambda^* V^{-1}(\mu-\mu_0 e) \\ \lambda^* &= \frac{r-\mu_0}{a-2\mu_0 c+\mu_0^2 b} =: \frac{r-\mu_0}{\Delta_R}. \end{aligned} \tag{7.33}$$

The locus of minimum variance portfolios is given by

$$\sigma_R(r) = \pm\frac{r-\mu_0}{\sqrt{\Delta_R}}\ . \tag{7.34}$$

Proof of Proposition 3.7.9.

Proof. The VaR_α for the level $1-\alpha$ satisfies by definition

$$P(R^\phi \leq -\mathrm{Var}_\alpha) \leq \alpha\ .$$

If P is a normal distribution with mean μ and variance σ the definition reads:

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{-\mathrm{VaR}_\alpha} e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \leq \alpha\ .$$

The change of coordinates $z=\frac{x-\mu}{\sigma}$ leads to

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{-\mathrm{VaR}_\alpha-\mu} e^{-\frac{1}{2}z^2}\sigma dz \leq \alpha\ ,$$

with σ the Jacobian, i.e.

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{\mathrm{VaR}_\alpha-\mu}{\sigma}} e^{-\frac{1}{2}z^2}dz \leq \alpha\ .$$

The upper integral limit depends on α,μ,σ. If we set the variance to 1 and the mean to zero, for a given α the critical value or VaR follows. For $\alpha=0.01$, i.e. a VaR on the 99 percent confidence level, a numerical solution of the equation

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{k_\alpha} e^{-\frac{1}{2}z^2}dz \leq 0.01$$

delivers the critical value $k_\alpha = -2.33$. Increasing the confidence level to 99.9 percent increases the critical level to $k_\alpha = -3.09$. We use this in our calculation. From

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\frac{\mathrm{VaR}_\alpha - \mu}{\sigma}} e^{-\frac{1}{2}z^2} dz \leq \alpha \ .$$

follows

$$-\frac{\mathrm{VaR}_\alpha - \mu}{\sigma} \leq k_\alpha \ .$$

In other words, the VaR_α under normality assumption is

$$-\mathrm{VaR}_\alpha \leq \sigma k_\alpha + \mu \ .$$

Since the VaR constraint binds, we have

$$-\mathrm{VaR}_\alpha = \sigma k_\alpha + \mu \ .$$

This is the VaR for a given period. If we calculate the VaR for a different period, the square-root-rule applies. That is, for a period T the VaR reads:

$$-\mathrm{VaR}_\alpha = (\sigma k_\alpha + \mu)\sqrt{T} \ .$$

□

7.8 Investments

7.8.1 CPPI

Proof of Proposition 3.4.1.
Starting with $c_t = V_t - F$ we get with the associated dynamics

$$dc_t = c_t \left(h dS_t / S_t - (h-1) r dt \right) \ .$$

Inserting the S-dynamics we get a geometric Brownian motion dynamic for the cushion with the solution given in the proposition. To prove the second claim we use the first one

$$dc_t / c_t = ((r + h(\mu - r))dt + \sigma h dW_t)$$

and the solution of the S-geometric Brownian motion. That is, solving the S-solution w.r.t. to W_t we get

$$W_t = \frac{1}{\sigma} \left(\ln(S_t/S_0) - (\mu - \frac{1}{2}\sigma^2)t \right) \ .$$

We insert this in the cushion dynamics solution

$$c_t = c_0 e^{(h(\mu - r) + r) - \frac{1}{2}h\sigma^2)t + h\sigma W_t} \ .$$

We get

$$c_t = c_0 \left(\frac{S_t}{S_0}\right)^h \exp\left((r - h(r - \frac{1}{2}\sigma^2) - \frac{1}{2}h^2\sigma^2)t\right) =: a_t S_t^h$$

with

$$a_t = c_0/S_0 e^{bt} \ , \ b = r - h(r - \frac{1}{2}\sigma^2) - \frac{1}{2}h\sigma^2 \ .$$

This proves the second claim.
We consider the cushion process in discrete time:

$$c^n_{t_{k+1}} = V^n_{t_{k+1}} - F_{t_{k+1}}.$$

$$\begin{aligned} c^n_{t_{k+1}} &= V^n_{t_{k+1}} - F_{t_{k+1}} \qquad (7.35)\\ &= \max\left(0, \frac{hc^n_{t_k}}{S_{t_k}}\right) S_{t_{k+1}} + \left(V^n_{t_k} - \max\left(0, \frac{hc^n_{t_k}}{S_{t_k}}\right) S_{t_k}\right) \frac{B_{t_{k+1}}}{B_{t_k}} - F_{t_{k+1}} \\ &= \begin{cases} F_{t_k}\frac{B_{t_{k+1}}}{B_{t_k}} + (V^n_{t_k} - F_{t_k})\left(h\frac{S_{t_{k+1}}}{S_{t_k}} - (h-1)\frac{B_{t_{k+1}}}{B_{t_k}}\right) - F_{t_{k+1}}, & \text{if } V^n_{t_k} - F_{t_k} > 0; \\ V^n_{t_k}\frac{B_{t_{k+1}}}{B_{t_k}}, & \text{else.} \end{cases} \end{aligned}$$

The risk free asset dynamics $F_{t_k}\frac{B_{t_{k+1}}}{B_{t_k}} = F_{t_{k+1}}$ implies

$$c^n_{t_{k+1}} = \begin{cases} (V^n_{t_k} - F_{t_k})\left(h\frac{S_{t_{k+1}}}{S_{t_k}} - (h-1)e^{rT/n}\right), & \text{if } V^n_{t_k} - F_{t_k} > 0; \\ (V^n_{t_k} - F_{t_k})e^{rT/n}, & \text{else} \ . \end{cases}$$

If we write t^* for the first discrete trading time where the cushion/investment account is negative the discrete time portfolio value at maturity T becomes

$$V^n_T = \begin{cases} V^n_{t^*}e^{r(T-t^*)}, & \text{if } t^* \le t_{n-1}; \\ G + (V^n_{t_{n-1}} - F_{t_{n-1}})\left(h\frac{S_{t_n}}{S_{t_{n-1}}} - (h-1)\frac{B_{t_n}}{B_{t_{n-1}}}\right), & \text{else} \ . \end{cases}$$

Summarizing, the discrete time cushion process can be written at a time $k+1$ as:

$$c^n_{t_{k+1}} = e^{r(t_{k+1} - \min(t^*, t_{k+1}))}(V^n_{T_0} - V_{t_0}) \prod_{j=1}^{\min(k+1, t^*)} \left(h\frac{S_{t_j}}{S_{t_{j-1}}} - (h-1)e^{rT/n}\right) \qquad (7.36)$$

This process converges toward the continuous time process.

7.8.2 Leveraged Negative Basis

The following notation is used.

$$
\begin{aligned}
t_k^j &= \text{Coupon date } t_k \text{ for Bond } j\text{, annual basis} \\
&\ \text{or premium payment date for CDS } j\text{, quarterly basis} \\
T^j &= \text{Maturity of bond (CDS) } j \\
N^j &= \text{Nominal amount of bond (CDS) } j \\
c(t_k^j) &= \text{Coupon of bond } j \text{ at date } t_k \\
\text{CDS} &= \text{CDS Premium} \\
n &= \text{Number of bonds (CDS) in the bond portfolio at date } t \\
w^j &= \text{Weight of bond } j \text{ at time } t = 0 \text{ in the bond portfolio} \\
\text{LE} &= \text{Leverage Unit, for example Euro } 50'000 \\
h &= \text{Leverage Factor} \\
\bar{\mathbf{n}}_{\mathbf{j}} &= \text{Negative Basis relative to bond issuer } j \text{ at issuance date of the certificate} \\
\bar{\mathbf{n}} &= \text{Negative Basis of all bond issuer at issuance date of the certificate} = \sum_{j=1}^{n} \omega^j \bar{\mathbf{n}}_{\mathbf{j}}
\end{aligned}
$$

$$
\begin{aligned}
\chi_Z &= \begin{cases} 1, & \text{if event } Z \text{ realizes;} \\ 0, & \text{else.} \end{cases} \\
D &= \text{for Default event} \\
T &= \text{for Trigger event} \\
\mathcal{E}(p) &= \text{6m Euribor starting in period } p \\
A &= \text{Financing spread} \\
D(t,T) &= \exp(-(T-t)(r+\text{ASW})) \ \text{ Discount factor of future Bond Cash Flows} \\
\overline{D}(t,T) &= \exp(-(T-t)(r+f)) \ \text{ Discount factor of past Bond Cash Flows} \\
& \quad\ r \text{ riskless interest rate, } f \text{ funding rate} \\
\text{ASW} &= \text{Asset Swap rate} \\
f &= A+\mathcal{E} \\
\tilde{t}^j &= \min(t_D^j, t_T^j) \quad \text{Default- vs. Trigger Event date comparison} \\
p &= \text{Semi-annual period for interest rate earnings resp. costs.} \\
R &= \text{Trigger Ratio} \\
Y &= \text{Year, 365 days (366 for leap years)} \\
s(p) &= \text{Balance in period } p,\ s(p) \leq 0 \\
E(p) &= \text{Earning in period } p \\
F(p) &= \text{Funding costs in period } p,\ F(p) \geq 0 \\
G &= \text{Annual fee for certificate} \\
\mathbf{p} &= \text{Participation rate at the negative basis}
\end{aligned}
$$

The portfolio $V(t)$, which represents both the underlying value and the value of the certificate, consists at each date t, $0 \leq t \leq T$, of three sub portfolios:

$$V(t) = V^B(t) + V^{\text{CDS}}(t) + V^{\text{Cash}}(t)$$

with $V^B(t)$ the bond portfolio, $V^{\text{CDS}}(t)$ the CDS portfolio and $V^{\text{Cash}}(t)$ the cash portfolio.

The value of the bond portfolio at time t reads:

$$V^B(t) = h \sum_{j=1}^{n} \underbrace{\chi_{\{\tilde{t}^j > t\}} \left(\sum_{t_k^j > t} w^j c(t_k^j) D(t, t_k^j) + \sum_{T^j > t} N^j w^j D(t, T^j) \right)}_{= w^j V(t)^{j,\text{Bond}}}$$

where

$$
\begin{aligned}
\chi_{\{\tilde{t}^j > t\}} &= \text{Default or Trigger event after time } t \\
\sum_{t_k^j > t} w^j c(t_k^j) D(t, t_k^j) &= \text{PV of coupons at } t \text{ for each bond } j \\
& \quad\ \text{with weights } w^j \text{ fixed at 0.} \\
\sum_{T^j > t} N^j w^j D(t, T^j) &= \text{PV of notional .}
\end{aligned}
$$

The value $V(t)^{j,\text{Bond}}$ is a mark-to-market value. All earnings of the bond portfolio up to time t are defined by:

$$\tilde{V}^B(t) = V^B(t) + \underbrace{\sum_{j=1}^{n} \sum_{t_k^j < t} \chi_{\{t_k^j > \tilde{t}^j\}} w^j c(t_k^j) \overline{D}(t, t_k^j)}_{=V(t)^{B,\text{Cash}}}$$

where

$$w^j c(t_k^j) \overline{D}(t, t_k^j) \quad = \quad \text{PV of all coupons} < t \text{ which did not defaulted} \quad .$$

The CDS portfolio is similarly split into a mark-to-market value at time t and past cash flows which enter the cash portfolio, i.e.

$$\begin{aligned} \tilde{V}^{\text{CDS}}(t) &= \underbrace{-h \sum_{j=1}^{n} \chi_{\{\tilde{t}^j > t\}} w^j N^j \sum_{t_k^j > t} \text{CDS} R_k^j D(t, t_k^j)}_{=V^{\text{CDS}}(t)} \\ &+ \underbrace{h \sum_{j=1}^{n} \chi_{\{\tilde{t}^j < t\}} w^j N^j \overline{D}(t, t_D^j)}_{=V^{\text{CDS, Cash}}(t)} \quad . \end{aligned}$$

At a default event, the CDS portfolio contribution is at market value. The cash portfolio consists of 3 further components: The loan part, the coupon part of the certificate and the trigger part in case of a trigger event. We refer to the main text for the discussion.

Chapter 8

References

- T. Adrian and M. Brunnermeier, CoVaR, Princeton Working Paper, 2011.
- T. Adrian and H. S. Shin. The Changing Nature of Financial Intermediation and the Financial Crisis of 2007-2009. Annual Review of Economics, 2(1):603–618, 2010.
- T. Adrian and H. S. Shin. Liquidity and Leverage, Journal of Financial Intermediation, 19 (3), 418-437, 2010.
- V. Agarwal, N. D. Daniel, and N. Y. Naik, Do Hedge Funds Manage Their Reported Returns?, Review of Financial Studies, forthcoming, 2011.
- F. Allen and D. Gale. Financial Contagion. Journal of Political Economy, 108(1):1–33, 2000.
- F. Allen and A. Babus. Networks in Finance. In P. Kleindorfer and J. Wind, editors, The Network Challenge, pages 367–382. Wharton, 2009.
- F. Allen and D. Gale, Optimal Security Design,, Review of Financial Studies 3(3), 229-263, 1988.
- F. Allen and D. Gale, Financial Innovation and Risk Sharing, MIT Press, Cambridge, MA, 1994.
- F. Allen, J. Barth and G. Yago, Fixing the Housing Market: Financial Innovations for the Future, Wharton School Publishing-Milken Institute Series on Financial Innovations, Upper Saddle River, NJ: Pearson Education, 2012.
- F. Allen and G. Yago, Financing the Futures. Market-Based Innovations for Growth. Wharton School of Publishing and Milken Institute, 2010.
- W. Allen and G. Wood, Defining and Achieving Financial Stability, Journal of Financial Stability, vol 2, issue 2,152–72, 2006.

- H. Amini, R. Cont, and A. Minca. Resilience to Contagion in Financial Networks. SSRN eLibrary, 2010.
- H. Amini, R. Cont, and A. Minca. Stress Testing the Resilience of Financial Networks. To appear in the International Journal of Theoretical and Applied Finance, 2011.
- E. Anderson, The Dynamics of Risk-Sensitive Allocations, Journal of Economic Theory 125 (2), 93-150, 2005.
- M. Arnsdorf, Central Counterparty Risk, Quantitative Research, JP Morgan, arXiv:1205.1533v1, 2012.
- P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, Coherent Measures of Risk, Mathematical Finance 9 (3): 203–228, 1999.
- V. Bala and S. Goyal. A non-Cooperative Model of Network Formation. Econometrica, 68:1181–1229, 2000.
- M.W. Baxter and A.J.O. Rennie, Financial Calculus, An Introduction to Derivative Pricing, Cambridge University Press, 2003.
- I. Ben-David, F. Franzoni, A. Landier and R. Moussawi, 2012, Do Hedge Funds Manipulate Stock Prices, Fisher College of Business Working Paper Series.
- S. Balder, M. Brandl and A. Mahayni, Effectiveness of CPPI Strategie under Discrete-Time Trading, Journal of Economic Dynamics and Control Volume 33, Issue 1, January 2009, Pages 204-220.
- Bank of England, Financial Stability Report, June, 2009.
- G. Beneplanc and J.-C. Rochet, Risk Management in Turbulent Times, Paul Woolley Research Initiative on Markets Disfunctionality at IDEI, Toulouse, 2010.
- K. Binmore, Rational Decisions, Princeton, NJ: Princeton University Press, 2009.
- F. Black, and M. Scholes, From Theory to a new Financial Product, Journal of Finance 29(2), 399-412, 1874.
- T. Björk, Arbitrage theory in continuous time, Oxford University Press, 2009.
- BlackRock, ETF landscape: Global Handbook Q1, 2011.
- G.F. Blundell and C. W. R. Ward, Property Portfolio Allocation: A Multi-Factor Model, Journal of Property Research 4 (2): 145–156, 1987.
- V. Bogan, Stock Market Participation and the Internet, Journal of Financial and Quantita- tive Analysis, Vol. 43, No. 01, pp. 191–211, 2008.

- C. Borio and M Drehman, Assessing the Risk of Banking Crises – Revisited , BIS Quarterly Review, March, 29-46, 2009.
- M. Boss, H. Elsinger, A. Lehar an M. Summer, The Network Topology of the Interbank Market, Quantitative Finance, 4, 677-684, 2004.
- D. van Bragt, M. Francke, B. Kramer and A. Pelsser, Risk-Neutral Valuation of Real Estate Derivatives, OFRC working paper series, Discussion Paper 10/2009 - 048, October, 2009
- M. Brenner and Y. Izhakian, Asset Prices and Ambiguity: Empirical Evidance, Stern School of Business, Finance Working Paper Series, FIN-11-10, 2011.
- D. Brigo, A. Pallavicini and R. Torrosetti, Credit Models and the Crisis, or: How I learned to stop worrying and love the CDOs, arXiv, 17 Feb 2010.
- D. Brigo and F. Mercurio, Interest-Rate Models - Theory and Practice. Springer Verlag, 2001.
- D. Brigo, A. Capponi, A. Pallavicini and V. Papatheodorou, Collateral Margining in Arbitrage-Free Counterparty Valuation Adjustment including Re-Hypotecation and Netting, Working Paper, King's College, London, 2011.
- R. Brian, F. Nielsen and D. Steffek, Portfolio of Risk Premia: A New Approach to Diversification, MSCI Barra Research Insights, 2009.
- J. Brogaard, High Frequency Trading and its Impact on Market quality, Northwestern University Kellogg School of Management Working Paper, September, 2010.
- M. Brunnermeier and S. Nagel, Hedge Funds and the Technology Bubble, Journal of Finance 59(5), 2013-2040, 2004.
- Cao, H. H., T. Wang, and H. H. Zhang, Model Uncertainty, Limited Market Participation, and Asset Prices, The Review of Financial Studies, Vol. 18, No. 4, pp. 1219–1251, 2005.
- P. Carr and L. Wu, Variance Risk Premiums, Review of Financial Studies, Vol. 22, No. 3, pp. 1311–1341, 2009.
- M. Chlistalla, High-Frequency Trading. Better Than its Reputation?, Deutsche Bank Research, February, 2011.
- R. Cifuentes. G. Ferrucci and H.Shin, Liquidity Risk and Contagion, Journal of the European Economic Association, 3, 556-566, 2005.
- J. Cochrane, Asset Pricing, Princeton University Press, 2005.
- G. Crawford and B. Sen, Derivatives for Decision Makers, John Wiley, New York, 1996

- R. Cont, A. Moussa and E. Santos, Network Structure and Systemic Risk in Banking System, Preprint, 2010.
- T. Coyler and M. Sebag-Montefiore, The State of European Bank Funding, Oliver Wyman, 2011.
- D. DiPasquale and W. Wheaton. Urban Economics and Real Estate Markets. Englewood Cliffs, NJ: Prentice Hall, 1996.
- V. De Miguel, L. Garlappi, F. Nogales and R. Uppal, A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms. Management Science, 55, 798-812, 2009.
- Y. Demyanyk and O. Van Hemert, Understanding the Subprime Mortgage Crisis, Review of Financial Studies, Volume 24, Issue 6, 1848-1880, 2011.
- E. Derman and I. Kani. The Volatility Smile and its Implied Tree. *Risk*, 7:32-39, 1994. (available on-line)
- E. Derman, I. Kani and N. Chriss. Implied Trinomial Trees of the Volatility Smile. Goldman Sachs Quantitative Strategies Research Notes, 1996. Risk, 7:32-39, 1996.
- B. Dupire. Pricing with a Smile. Risk, 7:18-20, 1994.
- J. Dow and S. R. d. C.Werlang, Uncertainty Aversion, Risk Aversion, and the Optimal Choice of Portfolio, Econometrica, Vol. 60, No. 1, pp. 197–204, 1992.
- I. Drechsler, Uncertainty, Time-Varying Fear, and Asset Prices, The Journal of Finance, Forthcoming, 2012.
- D. Duffie and H. Zhu, Does a Central Clearing Counterparty Reduce Counterparty Risk, The Review of Asset Pricing Studies, Volume 1, Issue 1, 74-95, 2012.
- L.G. Epstein and T. Wang, Intertemporal Asset Pricing Under Knightian Uncertainty, Econometrica, Vol. 62, No. 2, pp. 283–322, 1994.
- L. Eisenberg and T. H. Noe. Systemic Risk in Financial Systems. Management Science, 47(2):236–249, 2001.
- R. Elul, Welfare Effectgs of Financial Innovation in Incomplete Markets Economies with Several Consumption Goods, Journal of Economic Theory 65, 43- 78, 1995.
- B. Eraker, Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices, The Journal of Finance, Vol. 59, No. 3, pp. 1367–1404, 2004.
- H.-J. Jaeger, Tax Transparency as a Trend in International Taxation, Ernst&Young, Swiss Finance Institute Seminar, 2012.
- P. Erdõs and A. Rényi. On random graphs. I. Publ. Math. Debrecen, 6:290–297, 1959.

- F. Fabozzi, R. J. Shiller, and R. Tunaru, Hedging Real-Estate Risk, working paper 09-12, Yale International Center for Finance, 2009.
- E. Fama, Efficient Capital Markets: A Review of Theory and Empirical Work, Journal of Finance 25, 383–417, 1970.
- S. Feld, Why your Friends have More Friends than you do. Am. J. Sociol. 96, 1464–1477, 1991.
- D. Filipovic, Term-Structure Models: A Graduate Course, Springer Finance, Berlin, 2009.
- Financial Stability Board, Shadow Banking: Strengthening Oversight and Regulation, October, 2011.
- F. Franzoni, Empirical Asset Pricing, Lecture Notes, Swiss Finance Institute and University of Lugano, 2008.
- R. Frey and T. Schmidt, Vorlesungsskript Finanzmathematik I, University of Leipzig, 2006.
- P. Gai and S. Kapadia. Contagion in Financial Networks. Proceedings of the Royal Society A, 466(2120):2401–2423, 2010.
- Gai, P. and Kapadia, S. Liquidity Hoarding, Network Externalities, and Interbank Market Collapse. Proc. R. Soc. A 466, 2401–2423 (2010).
- G. Galati and R. Moessner, Macroprudential Policy - a Literature Review, BIS Working Papers No 337, 2010.
- J. Gatheral, The Volatility surface: a Practitioner's Guide, John Wiley and Sons, Inc., 2006.
- D. Geltner and J. Fisher, Pricing and Index Considerations in Commercial Real Estate Derivatives Journal of Portfolio Management Special Issue: Real Estate – 2007, pp. 1–21, 2007.
- D. Geltner and D.G. Miller, Commercial Real Estate Analysis and Investments, South-Western University, College Publishing, 2001.
- J. Geanuracos and B. Millar, The Power of Financial Innovation, Harper Business, New York, 1991.
- A. Giese, On the Pricing of Auto-Callable Equity Structures in the Presence of Stochastic Volatility and Stochastic Interest Rates, MathFinance Workshop, Germany 2006.

- G. Gigerenzer and G. Goldstein, Reasoning the Fast and Frugal Way: Models of Bounded Rationality, in Heuristics: The Foundations of Adaptive Behavior, eds Gigerenzer G., Hertwig R., Pachur T., editors. (New York: Oxford University Press;), 31-57, 2011.

- G. Gigerenzer and K. Volz, Coognitive Processes in Decisions Under Risk are not the Same as in Decisions Under Uncertainty, Front Neurosci., 6, 105, 2012.

- W.N. Goetzmann and A. Kumar, Equity Portfolio Diversification, Review of Finance, Vol. 12, No. 3, pp. 433–463, 2008.

- P.-M. Glauser, Value through Wealth Planning - Key trends in taxation of private investors, Swiss Finance Institute, 2011.

- P.-M. Glauser, Taxation, Advanced Wealth Management Executive Education, Swiss Finance Institute and Carnegie Mellon, Pittsburgh, 2012.

- J. Glover, Soaring Targe2t Imbalances Stoke German Risk Angst: Euro Credit, Bloomberg, April, 2012.

- The Origins of Value. The Financial Innovations that Created Modern Capital Markets, W. Goetzmann and K. Rouwenhorst (eds.), Oxford University Press, 2005.

- M. Gordy, A Risk-Factor Model Foundation for Ratings-Based Bank Capital Rules, Journal of Financial Intermediation 12(3), 199-232, 2003.

- G. Gorton and A. Metrick, Securitized Banking and the Run on Repo, forthcoming Journal of Financial Economics, 2011.

- Government Office for Science, The Future of Computer Trading in Financial Markets, Working Paper, UK Government's Foresight Project on The Future of Computer Trading in Financial Markets, 2011.

- H. Guo, Limited Stock Market Participation and Asset Prices in a Dynamic Economy, Journal of Financial and Quantitative Analysis, Vol. 39, No. 03, pp. 495–516, 2004.

- A. G. Haldane. Rethinking the Financial Networks. 2009.

- A. G. Haldane and R. M. May. Systemic Risk in Banking Ecosystems. Nature, 469:351–355, 2010.

- L. Hansen and T. Sargent, Robust Control and Model Uncertainty. American Economic Review 91 (2),60-66, 2001. University Press, 2008.

- L. Hansen and T. Sargent, Robustness, Princeton University Press, 2008.

- J. Hasanhodzic, A. W. Lo, and Emanuele Viola, Is It Real, or Is It Randomized?: A Financial Turing Test, MIT Working Papers, 2010.
- E. Heitfield, Using Guarantees and Credit Derivatives to Reduce Credit Risk Capital Requirements Under the New Basel Capital Accord, in Credit Derivatives: the Definitive Guide, J. Gregory (Ed.), Risk Books, 2003.
- B. Henderson and N. Pearson, The Dark Side of Financial Innovation, ssrn working paper series http://ssrn.com/abstract=1342654, 2009.
- J. Hull, Risk Management and Financial Institutions, Third edition, Wiley Finance, 2012.
- J. Hull and A. White, Valuing Credit Default Swap Options. Journal of Derivatives, 10, 3 (Spring 2003), 40-50, 2003.
- G. Huberman and Z. Wang, Arbitrage Pricing Theory, Federal Reserve Bank of New York Staff Reports, Staff Report no. 216, 2005.
- V. Izhakian, Ambiguity Measurement, Stern School of Business, Economics Working Paper Series, ECN-12-01, 2012.
- M. Jackson, A. Wolinsky. A strategic Model of Social and Economic Networks. J. Econ. Theory, 71(1996).
- S. Jaecklin, The Impact of International Regulation on Swiss Banking: Perfect Storm or Opportunity, Oliver Wyman, Swiss Finance Institute Ticino Seminar, 2012.
- F. Jamshidian, LIBOR and Swap Market Models and Measures. Finance and Stochastics, 1(4):293— 330, 1997.
- B. Jovanovic and A. Menkveld, Middlemen in Limit Order Markets, Working Paper, New York University, 2010.
- D. Kahneman, Thinking Fast and Slow. New York: Farrar, Straus Giroux, 2011.
- H.M. Kat, Structured Equity Derivatives: The Definitive Guide to Exotic Options and Structured Notes. Wiley Finance, 2008.
- C. Kindleberger and R. Aliber, Manias, Panics and Crashes: A History of Financial Crisis, CPI Antony Rowe, 2011.
- F. Knight, Risk, Uncertainty, and Profit, New York: Houghton Mifflin, 1921.
- Y.K. Kwok, Computational Methods for Pricing Structured Products, Lecture Notes, University of Hong Kong, 2011.
- D. Laibson, Golden Eggs and Hyperbolic Discounting, Quarterly Journal of Economics, Volume 112, Issue 2, 443-478, 1997.

- O. Ledoit and M. Wolf, Honey, I Shrunk the Sample Covariance Matrix, Journal of multivariate analysis, 2004.
- M. Leippold and P. Vanini, The Quantification of Operational Risk, Journal of Risk, Vol. 8, No. 1, Fall 2005.
- M. Leippold, Resampling and Robust Portfolio Optimization, Lecture Notes University of Zurich, 2010.
- Y. Leitner. Financial networks: Contagion, Commitment and Private Sector Bailouts. Journal of Finance, 60:2925–2953, 2005.
- J. Lintner, The Valuation of Risk Assets and the selection of Risky Investments in Stock Portfolios and Capital Budgets, Review of Economics and Statistics, 47 (1), 13-37, 1965.
- A. Lubloy (2006). Topology of the Hungarian Large-Value Transfer System. Occasional Papers 57, Magyar Nemzeti Bank, Budapest.
- R. M. May and N. Arinaminpathy. Systemic Risk: the Dynamics of Model Banking Systems, Journal of The Royal Society Interface, 7(46):823–838, 2010.
- R. L. McDonald, Derivatives Markets, Boston: Pearson International Edition, 2006.
- McKinsey&Company, National Regulations, Swiss Finance Institute, Ticino Seminar, 2012.
- R.C. Merton, Financial Innovation and Economic Performance, Journal of Applied Corporate Finance 4(4), 12-22, 1989.
- D. X. Li, On Default Correlation: A Copula Function Approach. Journal of Fixed Income, Vol. 9, No. 4, (March 2000), pp. 43-54, 2000.
- R.C. Merton, On the Application of the Continuous Time Theory of Finance to Financial Intermediation and Insurance, Geneva Papers on Risk and Insurance 14(July), 1989, 225-262.
- A. C. Minca, 2011, Modélisation Mathématique de la Contagion de Défault. Univ. Pierre et Marie Curies, Paris.
- M.H. Miller, Financial Innovation: The Last Twenty Years and the Next, Journal of Financial and Quantitative Analysis, Volume 21,Issue 04, December, 459-471, 1986.
- M.H. Miller, Financial Innovation and Market Volatility, Blackwell, Cambridge, MA, 1991.
- M.H. Miller, Debt and Taxes, Journal of Finance 32, 261-275, 1977.

- J. Mueller, Interbank Credit Lines as a Channel of Contagion, Journal of Financial Services Research, 29(1):37-60, 2006.
- B. Nathanael, The Basel Committees December 2009 Proposals on Counterparty Risk. Federal Reserve/FSA presentation at the International Workshop on Counter Party Risk Management and Application of CVA, Bank of Japan, June 14, 2010.
- M. Newman, Random Graphs as Models of Networks. In Handbook of graphs and networks (eds S. Bornholdt H. G. Schuster). Berlin, Germany: Wiley-VCH, 2003.
- S. N. Neftci, Principles of Financial Engineering, Academic Press Advanced Finance Series, 2nd edition, 2008.
- J. Mossin, Equilibrium in a Capital Asset Market, Econometrica, Vol. 34, No. 4, pp. 768–783, 1966.
- M. Padovani and P. Vanini, An Intergenerational Cross-Country Swap, Journal of Risk Finance, 11(5), 446-463, 2010.
- E. Pagnotta and T. Philippon, Competing on Speed, New York University Stern School of Business, working paper, 2012.
- C. Pirrong, The Economics of Clearing in Derivatives Markets: Netting, Asymmetric Information, and the Sharing of Default Risks Through a Central Counterparty, University of Houston, 2009.
- M. Pykhtin and A. Dev, Credit Risk in Asset Securitizations: an Analytical Model, Risk May 2003, 515-520, 2002.
- J. Quigley. Housing Economics and Public Policy, chapter Transaction costs and housing markets, pages 56-64. Oxford, Blackwell Publishing, 2002.
- V. Piterbarg, A Practitioner's Guide to Pricing and Hedging Callable LIBOR Exotics in Forward LIBOR Models. Working Paper. Bank of America, 1-57, 2003.
- M. Pykhtin, Model Foundations of the Basel III Standardised CVA Charge, Journal of Risk, July, 60-66, 2012.
- S. Ramaswamy, Market Structures and Systemic Risks of Exchange-Traded Funds, BIS Working Papers No 343. 2011.
- M. Rabin,Psychology and Economics, Journal of Economic Literature 36, 11-46, March 1998.
- M. Rabin and T. O'Donoghue, Doing it Now or Later, The American Economic Review Vol. 89, No. 1, pp. 103-124, 1999.
- R. Rebonato, Modern Pricing of Interest Rate Derivatives: The LIBOR Market Model and Beyond. Princeton University Press, 2002.

- Reuters, Update 1-Lehman Brothers Holdings is Focus of Grand Jury Probe, October 16, 2008.
- R. Roll, A Critique of the Asset Pricing Theory's Tests, Journal of Financial Economics 4: 129–176, 1977.
- D. Rosen and D. Saunders, Computing and Stress Testing Counter Party Credit Risk Capital, Counter Party Credit Risk, E. Canabarro (editor), Risk books, 2010.
- S. Ross, The Arbitrage Theory of Capital asset Pricing, Journal of Economic Theory 13, 341–60, 1976.
- S. Ross, Risk, Return and Arbitrage, Risk Return in Finance ed. I. Friend and J. Bicksler, Cambridge, Mass.: Ballinger, 1976
- A. Sadr, Interest Rate Swaps and Their Derivatives. A Practitioner's Guide, Wiley Finance, 2010.
- L. Savage, Foundations of Statistics, New York: Wiley, 1954.
- B. Scherer, Portfolio Construction and Risk Budgeting. Risk Books, 2002.
- A. Schinasi, Safeguarding Financial Stability, Theory and Practice, International Monetary Fund, 2004.
- D. Schmeidler, Subjective Probability and Expected Utility without Additivity, Econometrica, Vol. 57, No. 3, pp. 571–87, 1989.
- L. Seco, Part 1: Review of Hedge Funds and Structured Products, Lecture Notes, University of Toronto, RiskLab, 2010.
- W. F. Sharpe, Capital asset prices: A theory of market equilibrium under conditions of risk, Journal of Finance, 19 (3), 425-442, 1964.
- H. Shin, Risk and Liquidity in a System Context. Workshop on 'Accounting, risk management and prudential regulation', BIS Working Papers No 212, 2006.
- H. Shin, Risk and Liquidity in a System Context, Journal of Financial Intermediation, vol. 17, pp. 315– 29, 2008.
- H. Stein, Counterparty Risk, CVA, and Basel III, Columbia University, Financial Engineering Practitioners Seminar, 2012.
- 40. J. Stiglitz, Freefall: America, Free Markets, and the Sinking of the World Economy. New York: Norton, 2010.
- J. Syz and P. Vanini, Property Derivatives and Index-Linked Mortgages, Journal of Real Estate Finance and Economics, Vol. 36, No. 1, 2008.

- J. Syz and P. Vanini, Arbitrage Free Price Bounds for Property Derivatives, Journal of Real Estate Finance and Economics, Vol. 43, No. 3, 2011.
- A. Sussman, L. Tabb and R. Iati, US Equity High Frequency Trading: Strategies, Sizing and Market Structure, TABB Group, December, 2010.
- N. Taleb, The Black Swan. The Impact of the Highly Improbable. New York: Random House, 2010.
- C. Telmer, Valuation Theory: Single-Period State Price Model, Tepper School of Business, Carnegie Mellon University, 2007.
- J. L. Treynor, Toward a Theory of Market Value of Risky Assets. Unpublished manuscript. A final version was published in 1999, in Asset Pricing and Portfolio Performance: Models, Strategy and Performance Metrics. Robert A. Korajczyk (editor) London: Risk Books, pp. 15–22, 1962.
- F. Trojani and P. Vanini, A Note on Robustness in Merton's Model of Intertemporal Consumption and Portfolio Choice, Journal of Economic Dynamics and Control, Vol. 26, No. 3, 423-435, 2002.
- P. Tufano, Financial Innovation, The Handbook of the Economics of Finance (North Holland), edited by George Constantinides, Milt Harris and René Stulz, 2002.
- T. Ui, The Ambiguity Premium vs. the Risk Premium under Limited Market Participation, Review of Finance, Vol. 15, No. 2, pp. 245–275, 2011.
- J. Walmsley, The New Financial Instruments, John Wiley, New York, 1988.
- J. Wang, Arbitrage Pricing Theory, Lecture Notes, MIT, 2003.
- WWF and Credit Suisse, Transition to a Low Carbon Economy, 2011.
- U. Wystup, FX Options and Structured Products. Wiley Finance, 2006.

Index