INSURANCE MATHEMATICS

TEPPO A. RAKKOLAINEN

Foreword

This lecture notes were written for an advanced level course on Insurance Mathematics given at Åbo Akademi University during spring term 2010. While best efforts to correct all typos found during the lectures (many thanks for the students for pointing out the typos) have been made, the notes are still without doubt in very unpolished form.

The presentation relies most heavily on lecture notes [1] and textbook [6] with regard to most of Part 1 (dealing with mathematical finance), and on textbook [13] with regard to Parts 2 and 3 (dealing with classical life insurance mathematics and multiple state models, respectively). In Part 4, use has been made of several sources. References are not explicitly given in the text of the presentation, as this to me seems not to be really necessary considering the purpose of lecture notes.

To conclude this foreword, a thought on studying and attending lectures from the Devil:

"Habt Euch vorher wohl präpariert,
Paragraphos wohl einstudiert,
Damit Ihr nachher besser seht,
Dass er nichts sagt, als was im Buche steht."
(Mephistopheles in Goethe's *Faust*)

Turku, April 2010

Teppo Rakkolainen

Some minor corrections and additions were made in August 2013. *TR*

Contents

1. Some Financial Mathematics

1.1. **Motivation: On the Role of Investment in Insurance Business.**

In both life[1] and non-life insurance[2], insurers provide their customers with (usually partial) coverage for financial losses caused by potential adverse future events. In non-life insurance, an example of such an adverse event might be a fire causing damage to the insured party's residence. The insurer then covers the costs of repair works. Correspondingly, in a life insurance contract the unexpected death of the insured might trigger a series of pension payments to the surviving family members of the insured.

In all branches of insurance, the insurer receives *certain* (*deterministic*) payment or payments, insurance premiums[3], from the insured, in exchange for a contractual promise to cover financial losses caused by some specified *potential* – i.e. *uncertain* (*stochastic*) – future events. The amount received should cover the unknown losses arising during the contract period, the insurer's operating expenses, and additionally the insurer should obtain some profit. Hence the insurer receives payments in advance, before the covered events specified in the insurance contract have happened – often these events may not occur during the contract period at all. Thus the insurer ends up with significant temporary funds in its balance sheet, and these funds need to be invested profitably to generate investment returns, which can be used to offset the costs caused by incurred claims and the insurer's operational costs, and to increase the insurer's profit margin. These funds are liabilities for the insurer as they are meant to cover the losses arising from insured events to the insured parties. The insurer is expected to invest prudently, as it is, in a manner of speaking, investing other people's (the insured persons', in this case) money. The risk of the insurer not being able to meet its contractual obligations should remain on an acceptable level – in practice, this means that the insurer's assets should at all times with a high probability be sufficient to cover the liabilities. In other words, the insurer should have a sustainable solvency position. Moreover, the insurer needs to maintain sufficient liquidity to be able to pay the claim costs (which are not known in advance) as they realize. Hence it is not reasonable to invest too much into hard-to-realize illiquid assets even if they offer high returns.

From the previous considerations it should be obvious that in order to be able to generate good investment returns without excessive risk-taking and without compromising its liquidity and solvency position, any insurance undertaking must have some knowledge on financial markets and their functionality at its disposal. Good asset-liability management is also essential in avoiding situations where due to an asset–liability mismatch, say, values of assets plummet while liabilities' value remains unchanged or even increases. This leads to a weakening of the insurer's solvency position and possibly to insolvency.

Investment concerns are especially pronounced for life insurers, since their liabilities tend to have a long maturity: life and pension insurance contracts may have 30- or 50-year life spans. In a pension insurance contract, premiums may be paid by the insured for 30

---

[1] life insurance = livsförsäkring = henkivakuutus

[2] non-life / general / property and casualty (P & C) insurance = skadesförsäkring = vahinkovakuutus

[3] insurance premium = försäkringspremie = vakuutusmaksu

years, after which the insured receives pension payments for 20 years. Matching assets for such long-term liabilities are not easily found from financial markets, and since terms of the contract are usually fixed in the beginning, the insurer may not be able to adjust premiums to accommodate adverse developments. In contrast, non-life insurance contracts are usually renewed annually and the terms can be adjusted when contract is renewed. Hence the risk for an asset–liability mismatch is smaller.

Understanding of the workings of financial markets and the basic principles for valuation of financial assets is of fundamental importance for actuaries nowadays, since they have a responsibility not only to ensure appropriate and correct calculation of their company's liabilities, but also to ensure that the company's assets appropriately match these liabilities and that the company's investment strategy takes properly into account the requirements set by the nature of the liabilities. A total balance sheet approach and market consistent valuation (of both assets and liabilities) are also central principles in the Solvency II Directive creating a new solvency regime and regulatory framework for all life and non-life insurers in the E. U.

1.2. **Financial Markets and Market Participants.** Financial markets bring providers of capital together with the users of capital. Their function is to facilitate channeling of funds from savers to investors. In this role the markets complement the financial intermediaries (banks, insurance companies etc.) and compete with these institutions. On the other hand, financial institutions are also important players in financial markets.

Some major reasons for the existence of financial markets are the following:

(1) They enable consumption transfer over time;
(2) They enable risk sharing, hedging and risk transfer among market participants;
(3) They enable conversions of wealth;
(4) They produce information and improve allocation of resources to most productive uses.

Financial markets also provide fascinating opportunities for gamblers.

An important concept central to much of the theory of mathematical finance is *market efficiency*. In efficient markets, the market prices reflect all investors' expectations given the set of available information. Prices of all assets equal their investment value at all times, and forecasting market movements or seeking undervalued assets based on available information is futile: you cannot beat the market, which incorporates any new information immediately into prices. Put differently, investors cannot expect to make abnormal profits by using the available information to formulate buying and selling decisions: they can only expect a normal rate of return on their investments.

Market efficiency can be classified according to what information exactly is considered to be fully reflected in market prices. In *weak form efficiency*, past price information cannot predict future prices; in *semi strong form efficiency*, prices contain all publicly available information; and in *strong form efficiency* prices contain all information, including insider information.

An important aspect of markets is their *liquidity* (or lack of it). In a liquid market, an investor wishing to sell a security will always easily find a counterparty willing to buy the security at market price - an individual transaction has no impact on the price. Characteristic for deeply liquid markets is an abundance of ready and willing buyers and sellers so that market prices can be determined by demand and supply in the marketplace. In illiquid markets, this price discovery process does not operate due to a lack of ready and willing buyers or sellers. In this sense illiquid markets cannot be efficient as the market pricing mechanism does not work. Fundamentally illiquidity is due to investors' uncertainty about the true value of a security. Importance of liquidity (it has been called "oxygen for a functioning market", reflecting the fact that while it is available it is not really noticed, but a lack of it is immediately observable and has dire consequences) has been reinforced by the recent financial crisis begun by the collapse of the U. S. subprime mortgage market in 2007-2008.

One of the central areas of interest in mathematical finance is *valuation* or *pricing* of assets. In this presentation we will later on consider valuation of some specific financial assets in more detail. At this point we only point out that pricing methodologies can broadly be classified into two groups: principles based on *discounted cash flow* and principles based on *arbitrage-free valuation.* The term *arbitrage* refers to an investment opportunity providing risk-free (excess) returns, that is, certain profit without any risk – a money making automaton. Such an opportunity implies that there is a misalignment of prices – someone is providing the "free"money without being compensated for this. In deeply liquid efficient markets, no arbitrage opportunities should exist, as such opportunities are constantly sought out by investors and taken advantage of as they appear, whereupon the price of the security in question goes up with increased demand until the price misalignment which gave rise to the arbitrage has completely disappeared. In liquid efficient markets, this process is almost instantaneous as large sophisticated investors monitor markets continuously seeking arbitrage opportunities.

The most important submarkets of financial markets are

***money market*:** where corporations and governments borrow short term by issuing securities with maturity less than a year;

***bond market*:** where corporations and governments borrow long term by issuing bonds;

***stock market*:** where corporations raise capital by issuing shares or stocks, certificates conveying certain rights to their holders, and these shares are traded between investors in so-called secondary markets; and

***derivative market*:** where financial derivatives, that is securities whose value is a function of the values of some other securities (so-called underlying securities), are traded, the most typical examples being options, forwards, futures and swaps.

In all the submarkets the ownership of securities can be transferred between investors after issuance through so-called secondary markets; however, the liquidity of secondary markets can vary considerably depending on the security. While stocks traded on an organized exchange are typically highly liquid, bonds with very long maturities or exotic derivatives may be quite illiquid. Additionally, liquidity of a market is not constant in time – during periods of financial stress, markets previously considered liquid may "dry up" and become illiquid. Again, the subprime mortgage crisis provides an example: mortgage-backed securities with high credit ratings were quite liquid (there were willing buyers) as long as the credit rating was believed to reflect the true value of the securities, but as it became clear that ratings were way out of line with actual reality, the investors' uncertainty considering the true value of these securities increased to the point causing almost total illiquidity – there were virtually no willing buyers at all.

Market participants can broadly be classified into three groups: *speculators* try to obtain profits by making bets on the future direction of markets; *arbitrageurs* try to obtain profits by taking advantage of inconsistencies in prices between different securities ; while *hedgers* seek to reduce their risk exposures by taking positions on the market. Speculators and arbitrageurs play an important role in enhancing liquidity and efficiency of the markets. They are the ones constantly monitoring the prices and processing information; their investment decisions then influence supply and demand which determine the market prices.

1.3. **Interest Rates.** In a loan agreement, one party (the *lender*[4]) loans a specified amount of money (the *principal*[5] of the loan) to another party (the *borrower*[6]) for a specified time period, during which the borrower repays the principal and makes some additional compensation payments to the lender. *Interest*[7] is the compensation required by the lenders for lending funds to borrowers. To make this precise:

Definition 1.2.1: Suppose that a sum of B_0 euros is loaned by a lender to a borrower today, and it is agreed that the borrower pays the lender $B_1 > B_0$ euros one year from now. Then $B_1 - B_0$ is the *interest* on this loan paid by the borrower. The (one-year or *annual*) *interest rate* on the loan is

$$r := \frac{B_1 - B_0}{B_0}$$

and the corresponding *accumulation factor*[8] is

$$R := 1 + r.$$

More generally, r (resp. R) is an annual interest rate (resp. accumulation factor), if it holds that

$$B_t = (1 + r)^t B_0 = R^t B_0,$$

where the amount B_t is the value of the loan (the amount borrower owes to lender) at time t.

Interest rates are usually quoted in nominal terms. The *real interest rate* is determined as the difference of the nominal interest rate and the inflation rate. In this presentation we will mostly deal with nominal quantities, interest rates or other. It should be observed,

[4]lender = långivare = lainanantaja
[5]principal = kapital = pääoma
[6]borrower = låntagare = lainanottaja
[7]interest = ränta = korko
[8]accumulation factor = kapitaliseringsfaktor = korkotekijä

however, that inflation plays a very significant role in many branches of insurance, e.g. benefits of pension funds are often linked to consumer price indices or other quantities closely related to inflation.

There are several reasons for why interest is required by the lenders:

(1) The lender foregoes the possibility to increase his/her consumption by the amount he/she borrows to someone. In an inflationary environment, the purchasing power of money erodes as time passes, so without additional compensation the lender would suffer a loss of purchasing power since the nominal amount of money he/she receives at the end of the loan period is worth less in real terms than the nominal amount that he/she lent. Moreover, delaying consumption also means taking on the risk of not being able to consume the amount delayed later.
(2) The lender could have invested the principal of the loan to some other investment and would then have received corresponding returns from this investment: not obtaining these returns is a cost which should be compensated by the interest of the loan. Another way of saying this is that the loan is an investment for the lender, and interest payments by the borrower are the returns on this investment.
(3) There is a risk that the borrower *defaults*: i.e. is not able to pay back a part or all of his/her loan or interest payments. Hence a loan is a risky investment, and the lender requires an additional compensation for the default risk he/she takes on when loaning money.

Returning briefly to Definition 1.2.1, observe that from the lender's point of view, r is the *return* and R the *total return* of the loan investment. For the lender, the loan to the borrower is a financial asset generating positive returns (gains) in the form of interest income. For the borrower, the loan is a financial liability generating negative returns (losses) in the form of interest payments.

The above list can be summarized succinctly by the *time value of money*: one euro received today is worth more than one euro received one year from today. In financial mathematics, this is formalized by the concept of *present value*[9]: the nominal value of a cash flow received in the future is *discounted*[10] to the present time with the appropriate interest rate (*discount rate*). Discounting is done by multiplying the nominal value of the cash flow with a *discount factor*[11] defined (as a function of the appropriate interest rate r) as

$$v := \frac{1}{1+r} = \frac{1}{R}. \tag{1}$$

Thus, if the prevailing one-year risk-free interest rate is r %, the present value of a cash flow consisting of K euros a year from today is $vK = \frac{K}{1+r}$. This reflects the fact that with an initial investment of $\frac{K}{1+r}$ euros to an asset yielding the risk-free return r, any investor can (without any risk, i.e. with certainty) obtain K euros a year from today. The converse

[9] present value = nuvärde = nykyarvo
[10] discounting = diskontering = diskonttaus
[11] discount factor = diskonteringsfaktor = diskonttaustekijä

operation to discounting is called *accumulation*[12] and the discount factor is the inverse of accumulation factor defined in Definition 1.2.1. More generally, given an appropriate discount rate r, the discount factor for a cash flow of K euros which is realized at time t is $v^t K = \frac{K}{(1+r)^t}$.

Example: Suppose that an investor expects to receive a cash flow consisting of $K = 1000$ euros each year at the end of the year for the next 4 years and that the risk-free interest rate is constant $r = 0.03$. The present value of this cash flow is

$$\sum_{t=1}^{4} \frac{K}{(1+r)^t} = K \sum_{t=1}^{4} \frac{1}{(1+r)^t} = 1000 \cdot \left(\frac{1}{1.03} + \frac{1}{1.03^2} + \frac{1}{1.03^3} + \frac{1}{1.03^4} \right) = 3717.10.$$

In practice, there are several different interest rates in financial markets. As interest rates reflect the return on investment required by the lenders, it is clear that interest rates vary depending on the risk characteristics of the loan, such as the maturity time of the loan and creditworthiness of the borrower. To facilitate comparison between interest rates of different maturities, rates are usually expressed as *annualized* rates.

Example: Suppose that an investor has two alternatives: with an investment of 100 euros, the investor can obtain 120 euros in two years. The total return of this investment is $120/100 = 120\%$, and hence the relative return is 20 %. Alternatively, the investor can obtain 110 euros in one year. The total return of this second investment opportunity is $110/100 = 110\%$, and the relative return thus 10%. However, these returns are not directly comparable, since they have different maturities. To compare the attractiveness of these investments, we must convert the returns to annual returns. The annualized return r_2 of the two-year investment is the solution of equation $100(1 + r_2)^2 = 120$, which is easily calculated to be $r_2 = 0.0954 = 9.54\%$. This can be directly compared to the return of the second investment (which is already an annual return) and we see that the second opportunity is better, as its annual return $10\% > 9.54\%$, the annual return of the first opportunity.

Annualized interest rates or returns are indicated by adding '(p. a.)' (short for Latin *per annum*) after the quantity (e.g. 5 % (p. a.)).

So far we have considered so-called compound interest. *Compound interest*[13] means that the accrued interest is calculated periodically and added to the principal: on the next period, interest is earned not only on the original principal but also on the interest of previous periods. Of importance is the *compounding frequency*: that is, how often is the accrued interest added to principal. Suppose the effective interest rate is r % (p. a.) . If interest is added to principal m times a year, an initial principal of V_0 euros will grow in n years to

$$V(n) = (1+r)^n V_0 = \left(1 + \frac{r^{(m)}}{m}\right)^{nm} V_0 \tag{2}$$

[12] accumulation = kapitalisering, diskontering framåt= korkouttaminen, prolongointi

[13] compound interest = sammansatt ränta = koronkorko, yhdistetty korko

euros, where

$$r^{(m)} = m\left[(1+r)^{\frac{1}{m}} - 1\right] \tag{3}$$

is the nominal interest rate compounded m times per year equivalent to effective annual rate r. It is an easy exercise to show that $r^{(m)} < r$ for $m > 1$. Interest rate $r^{(m)}$ is a simple interest rate. With *simple interest*[14], interest is earned only on the principal, linearly with time. In general, if simple interest of r % (p. a.) is paid to a principal of V_0 euros, in t years the invested capital increases to

$$V(t) = (1 + rt)V_0. \tag{4}$$

Example: A 6-month deposit of K euros with a simple interest rate r % (p. a.) grows during the first month to $(1+\frac{r}{12})K$, during the first three months to $(1+\frac{r}{4})K$ and by maturity 6 months to $(1+\frac{r}{2})K$.

Correspondingly, we define the discount factor and the nominal accumulation factor for mth fraction of a year as

$$v^{(m)} := \frac{1}{1+\frac{r^{(m)}}{m}} = (1+r)^{-\frac{1}{m}} = v^{\frac{1}{m}} \tag{5}$$

and

$$R^{(m)} := 1 + \frac{r^{(m)}}{m} = (1+r)^{\frac{1}{m}} = R^{\frac{1}{m}} \tag{6}$$

Suppose now that we let the compounding frequency $m \to \infty$ and define

$$\delta := \lim_{m\to\infty} r^{(m)} = \lim_{m\to\infty} \frac{(1+r)^{\frac{1}{m}} - (1+r)^0}{\frac{1}{m}}; \tag{7}$$

from which we see that δ is the derivative of function $(1+r)^x$ at point $x = 0$. Hence,

$$\delta = \ln(1+r) \Leftrightarrow e^{\delta} = 1 + r = R. \tag{8}$$

Hence the final value of an n-year investment of initial capital V_0 earning interest r % (p. a.) with *continuous compounding*, i.e. when accrued interest is immediately (continuously) added to principal, is

$$\tilde{V}_n = V_0 e^{\delta n}. \tag{9}$$

Thus we see that when interest rate is r (p. a.) with continuous compounding, then the discount factor

$$v = e^{-\delta}. \tag{10}$$

Here δ is sometimes called the *force of interest*[15].

A hybrid of simple and compound interest is the interest rate used by Finnish banks (so-called ”pankkikorko”), where complete years are calculated using compound interest

[14] simple interest = enkel ränta = yksinkertainen korko

[15] force of interest = ränteintensitet = korkoutuvuus

but a fraction of a year is interpolated linearly with simple interest. For an investment earning this interest, an initial capital V_0 increases in t years to

$$V(t) = (1 + r(t - \lfloor t \rfloor))\,(1 + r)^{\lfloor t \rfloor} V_0, \tag{11}$$

where $\lfloor x \rfloor$ is the largest integer less than or equal to x. Observe that there are different *day count basis conventions* with regard to how the value of $t - \lfloor t \rfloor$ is transformed from a number of calendar days into a real number; in the following example we use the *English* or *actual/365* convention.

Example: Suppose that on 15.1.2010 a bank loans 150000 euros with annual interest rate $r_B = 0.03$ and that the principal is repaid on 30.6.2011 with no intermediate repayments. We calculate the bank's expected interest income from this loan. Loan period T is one full year and $16 + 28 + 31 + 30 + 31 + 30 = 166$ days, so that $T - \lfloor T \rfloor = \frac{166}{365} = 0.45479$, and hence by equation (11) the bank's expected interest income is

$$V(T) - V(0) = \left(1 + 0.03 \cdot \frac{166}{365}\right) \cdot 1.03 \cdot 150000 - 150000 = 6607.95.$$

Above we have considered the force of interest as constant in time. This can be generalized to accommodate variation in time by allowing δ to be a non-constant function of time.

Definition 1.2.2: Let $\delta : I \to \mathbb{R}$ be a piecewise continuous function on an interval $I \subset \mathbb{R}_+$. Function δ is a *force of interest*, if there exists a piecewise differentiable, continuous function $V : I \to \mathbb{R}$ such that

$$V^{'}(t) = \delta(t)V(t) \tag{12}$$

for all $t \in I \setminus N$, where N is the set of points of discontinuity of δ and points of nondifferentiability of V.

V is called a *provision*.

It is a simple exercise to show that if δ is a force of interest on $[a, b]$, then the provision at time $t \in [a, b]$ is given by

$$V(t) = \exp\left(\int_a^t \delta(u)du\right) \cdot V(a). \tag{13}$$

Suppose that δ is a force of interest and that $u = t_0 < t_1 < \ldots < t_{n-1} < t_n = t$. By virtue of the additivity of integrals, the provision can in this case be written as

$$V(t) = V(u)\exp\left(\int_u^{t_1} \delta(s)ds\right)\exp\left(\int_{t_1}^{t_2} \delta(s)ds\right)\ldots\exp\left(\int_{t_{n-1}}^{t} \delta(s)ds\right). \tag{14}$$

If δ is a constant δ_i on each interval (t_i, t_{i+1}), then this simplifies to

$$V(t) = V(u)\exp\left(\sum_{k=0}^{n-1} \delta_k(t_{k+1} - t_k)\right) = V(u)\prod_{k=0}^{n-1}(1 + r_k).^{t_{k+1}-t_k}, \tag{15}$$

where $r_k = e^{\delta_k} - 1$ is the annual interest rate corresponding to force of interest δ_k. The present value of provision $V(t)$ at time $u < t$ is

$$V(u) = V(t) \exp\left(-\int_u^t \delta(s)ds\right). \tag{16}$$

1.4. **Spot and Forward Rates.** Loans paying no intermediate interest or amortization payments, but only one single payment consisting of principal and interest at maturity time are called *zero-coupon loans*. Interest rate required by the lenders on a (default-free) zero-coupon loan with maturity $T > 0$ years beginning from today is called the T-year *spot rate*. Hence spot rates are determined by prices of zero-coupon securities: if the future price of a zero-coupon security at time T is P_T, and the current price is P_0, then the T-year spot rate

$$s_T = \left(\frac{P_T}{P_0}\right)^{\frac{1}{T}} - 1. \tag{17}$$

In other words, knowing the prices of zero-coupon loans for a set of maturities is equivalent to knowing the spot rates for these maturities.

Example: Suppose that a borrower can obtain a loan of $P_0 = 100000$ euros by agreeing to pay back $P_2 = 104040$ euros after 2 years. Then the two-year spot rate $s_2 = \left(\frac{104040}{100000}\right)^{1/2} - 1 = 0.02$.

There are also loan contracts which begin at some specified time in the future, say $S > 0$ years from now. Interest rate required by lenders on such a loan (with maturity T) is called the *forward rate*[16] from $T - S$ to T. If we assume that the markets are *arbitrage-free* (meaning that there are no possibilities to make certain excess returns over and above the risk-free rate of return), then forward rates are determined by spot rates, since to avoid arbitrage opportunities we must have

$$(1 + s_T)^T V_0 = (1 + s_{T-S})^{T-S}(1 + f_{T-S,T})^S V_0 \Leftrightarrow f_{T-S,T} = \left(\frac{(1 + s_T)^T}{(1 + s_{T-S})^{T-S}}\right)^{1/S} - 1,$$

where s_U is the U-year spot rate and $f_{V,W}$ is the forward rate for period $[V, W]$. That is, our return must be equal irrespective of whether we make a T-year investment of V_0 euros on the spot market today, or make a $T - S$-year investment of V_0 euros on the spot market plus enter a (forward) contract where we agree at time $T - S$ to invest $(1 + s_{T-S})^{T-S} V_0$ euros for S years at current forward rate for that period.

Example: Suppose that we know the one-year and two-year spot rates $s_1 = 0.012$ and $s_2 = 0.021$. Then the forward rate for an investment that will be made one year from now for a period of a year must be

$$f_{12} = \frac{(1 + s_2)^2}{1 + s_1} - 1 = \frac{1.021^2}{1.012} - 1 = 0.030$$

in order to avoid arbitrage.

[16] forward rate = terminränta = termiinikorko

In general, given the spot rates s_{T_i} for an increasing sequence of maturity times T_i, $i = 0, 1, 2, \ldots, N$, the forward rates f_{T_i,T_j} , $T_i < T_j$, can be calculated from

$$(1 + f_{T_i,T_j})^{T_j - T_i} = \frac{(1 + s_{T_j})^{T_j}}{(1 + s_{T_i})^{T_i}}. \tag{18}$$

In the continuous time framework, we denote the continuous spot rate at time t with maturity T by $s(t, T)$. For different maturity times $T_i \geq 0$, where $T_i < T_{i+1}$, the market price of a zero-coupon unit security is

$$P(t, T_i) = e^{-(T_i - t)s(t,T_i)}.$$

Now, denoting the continuously compounded forward rate available at time t for borrowing at time T_i and repaying at time T_j by $f(t, T_i, T_j)$, the fundamental arbitrage relation takes the form

$$e^{(T_j - t)s(t,T_j)} = e^{(T_i - t)s(t,T_i)} e^{(T_j - T_i)f(t,T_i,T_j)}.$$

This implies that

$$f(t, T_i, T_j) = -\frac{\ln P(t, T_j) - \ln P(t, T_i)}{T_j - T_i}. \tag{19}$$

The instantaneous forward rate at time t for borrowing at time T_i is defined as the limit

$$f(t, T_i) = \lim_{T_j \downarrow T_i} f(t, T_i, T_j) = -\frac{\partial}{\partial T_i} [\ln P(t, T_i)]$$

Integrating both sides and denoting $T_i = T$ yields

$$\int_t^T f(t, u)du = -\int_t^T \frac{\partial}{\partial u} [\ln P(t, u)]\, du \Leftrightarrow \ln P(t, T) = -\int_t^T f(t, u)du.$$

Comparison with $P(t, T) = e^{-(T-t)s(t,T)}$ shows that the continuous spot rate has the representation

$$s(t, T) = \frac{1}{T - t} \int_t^T f(t, u)du$$

in terms of the instantaneous forward rate. The continuous discount factor at time t with maturity T and time to maturity $T - t$ is hence

$$v(t, T) = e^{-s(t;T)(T-t)} = e^{-\int_t^T f(t,u)du}.$$

1.5. **Term Structure of Interest Rates.** Spot rates for different maturities $\{r_U | U > 0\}$ form the *term structure of interest rates*[17], i.e. interest rate as a function of maturity. In practice, the short end of this term structure curve is directly observable in the market (money market securities are usually zero-coupon instruments with no intermediate payments before maturity) but spot rates for the long-term market are not immediately observable as securities make intermediate interest payments (coupons). Assets used in deriving the term structure should belong to the same risk class, i.e. there are different

[17] term structure of interest rates = räntekurva = korkokäyrä, korkojen aikarakenne

term structures for government securities (risk-free or default-free term structure) and corporate securities of differing default risk (for example, one term structure for AAA rated corporations and another term structure for BBB rated corporations).

Example: Suppose that the prices of 1, 2 and 3-year zero-coupon loans with unit principal are $P_1 = 0.9850$, $P_2 = 0.9525$ and $P_3 = 0.9285$. Then the corresponding spot rates are

$$s_1 = \left(\frac{1}{0.9850}\right)^{1/1} - 1 = 0.015,\ s_2 = \left(\frac{1}{0.9525}\right)^{1/2} - 1 = 0.025,\ s_3 = \left(\frac{1}{0.9685}\right)^{1/3} - 1 = 0.035.$$

In the above example the term structure is upward sloping, i.e. spot rates for longer maturities are higher than spot rates for shorter maturities. This is usually the case, but the term structure may occasionally also be hump-shaped or downward sloping. To understand why this is so, recall that any interest rate reflects the required investment return of an asset. This required return can be decomposed into the following components:

(1) the required real return for the investment horizon, which can be considered to consist of the one-year *real risk-free interest rate* and a *term premium* with respect to the one-year rate, which depends on the investment horizon;
(2) expected inflation over the investment horizon; and
(3) a risk premium reflecting the additional return needed to compensate investor for additional risks such as default risk and illiquidity risk.

Real return reflects how much the investor wants his purchasing power to increase in compensation for delaying consumption. The longer the delay, the larger the compensation, and hence the term premium increases with investment horizon. This is one reason for the usual upward sloping term structure.

In an inflationary environment, investors require a compensation equal to the expected inflation over the investment horizon to shield them from the erosion of purchasing power. If expected inflation is constant, this leads to an upward shift of the whole term structure; however, should inflation expectations be lower in the long run than in the short term, this may cause the term structure to have a humped or downward-sloping shape.

If an asset is subject to default risk (or any other additional risk), a risk premium on top of the risk-free real return and expected inflation is required to make the asset attractive to investors. Magnitude of the default risk premium depends on the creditworthiness of the asset's issuer and also on the general market sentiment (*risk appetite*). Risk premiums for a specific issuer may also differ depending on the investment horizon. In the aftermath (?) of the recent financial crisis the existence and significance of an illiquidity premium component in the risk premium have come into focus. One line of thought is that such a premium tends to be negligible during "normal"times but can very rapidly increase in periods of financial stress.

1.6. **Annuities.**

Definition 1.4.1: An *annuity*[18] is a cash flow consisting of annual payments of equal magnitude. If the annual payment is equal to one unit, we speak of *unit annuity*.

If the first payment occurs at time 0 (at beginning of period), the annuity is called a *annuity due*[19] and the present value of a n-year unit annuity is given by

$$\ddot{a}_{\overline{n}|} := 1 + v + v^2 + \ldots + v^{n-1} = \sum_{i=0}^{n-1} v^i, \tag{20}$$

where $v = \frac{1}{1+r} = \frac{1}{R}$ is the discount factor. Corresponding *accrued provision* is

$$\ddot{s}_{\overline{n}|} := R + R^2 + \ldots + R^n = \sum_{i=1}^{n} R^i. \tag{21}$$

Annuities due are typically encountered in insurance contracts: the first premium is paid before the insurance cover is in force.

If the first payment occurs at time 1 (at end of period), we speak of *immediate annuity*[20], and the present value of a n-year unit annuity is given by

$$a_{\overline{n}|} := v + v^2 + \ldots + v^n = \sum_{i=1}^{n} v^i. \tag{22}$$

Corresponding *accrued provision* is

$$s_{\overline{n}|} := 1 + R + R^2 + \ldots + R^{n-1} = \sum_{i=0}^{n-1} R^i \tag{23}$$

Immediate annuities are encountered in bank loans: amortization and interest payments are usually paid at the end of the payment period.

An annuity with infinite duration is called a *perpetuity*[21]. Present values of perpetuities are obtained as limits of previous expressions as $n \to \infty$. Observe that the present value and accrued provision for immediate annuity can be obtained from the present value and accrued provision for annuity due via discounting:

$$a_{\overline{n}|} = v \cdot \ddot{a}_{\overline{n}|} \text{ and } s_{\overline{n}|} = v \cdot \ddot{s}_{\overline{n}|}. \tag{24}$$

Conversely, these quantities for annuity due can be obtained from their counterparts for immediate annuity via prolongation:

$$\ddot{a}_{\overline{n}|} = R \cdot a_{\overline{n}|} \text{ and } \ddot{s}_{\overline{n}|} = R \cdot s_{\overline{n}|}. \tag{25}$$

For $r \neq 0$, we have the following expressions for the present values and accrued provisions of annuities:

(i) : $\ddot{a}_{\overline{n}|} = \frac{1-v^n}{r \cdot v}$ and $\ddot{a}_{\overline{\infty}|} = \frac{1}{r \cdot v}$, for $r > 0$;

[18] annuity = annuitet, tidsränta = annuiteetti, aikakorko

[19] annuity due = annuitet på förhand = etukäteinen annuiteetti

[20] immediate annuity = annuitet på efterhand = jälkikäteinen annuiteetti

[21] perpetuity = päättymätön aikakorko/annuiteetti = oändlig tidsränta/annuitet

(ii) : $a_{\overline{n}|} = \frac{1-v^n}{r}$ and $a_{\overline{\infty}|} = \frac{1}{r}$, for $r > 0$;
(iii): $\ddot{s}_{\overline{n}|} = \frac{R^n-1}{r \cdot v}$;
(iv): $s_{\overline{n}|} = \frac{R^n-1}{r}$.

To prove the first part of (i), we use the formula for the value of a geometric sum:

$$\ddot{a}_{\overline{n}|} = \sum_{i=0}^{n-1} v^i = \frac{1-v^n}{1-v} = \frac{1-v^n}{1-\frac{1}{1+r}} = \frac{1-v^n}{\frac{1+r-1}{1+r}} = \frac{1-v^n}{r \cdot \frac{1}{1+r}} = \frac{1-v^n}{r \cdot v}.$$

The second part of (i) follows now easily by letting $n \to \infty$, since for $r > 0$ we have $|v| < 1$. To prove assertions (ii)-(iv) is left as an exercise.

In reality, payments are often made more frequently than annually. To accommodate this, we generalize the definition of annuity as follows.

Definition 1.4.2: An *annuity payable m times a year* is a cash flow consisting of payments of equal magnitude made at equal intervals m times a year . If the individual payment is equal to $\frac{1}{m}$, we speak of a *unit annuity.*

The *present value* and *accrued provision* of an *n-year unit annuity due payable m times a year* are

$$\ddot{a}_{\overline{n}|}^{(m)} := \frac{1}{m}\left(1 + v^{(m)} + {v^{(m)}}^2 + \ldots + {v^{(m)}}^{nm-1}\right) \tag{26}$$

and

$$\ddot{s}_{\overline{n}|}^{(m)} := \frac{1}{m}\left(R^{(m)} + {R^{(m)}}^2 + \ldots + {R^{(m)}}^{nm}\right). \tag{27}$$

Similarly, the *present value* and *accrued provision* of an *immediate n-year unit annuity payable m times a year* are

$$a_{\overline{n}|}^{(m)} := \frac{1}{m}\left(v^{(m)} + {v^{(m)}}^2 + \ldots + {v^{(m)}}^{nm}\right) \tag{28}$$

and

$$s_{\overline{n}|}^{(m)} := \frac{1}{m}\left(1 + R^{(m)} + {R^{(m)}}^2 + \ldots + {R^{(m)}}^{nm-1}\right). \tag{29}$$

As previously, the corresponding values for unit perpetuities payable m times a year are obtained by letting $n \to \infty$ in the expressions above.

If $r^{(m)} \neq 0$, then for unit annuities payable m times a year, the present values and accrued provisions are

(i) : $\ddot{a}_{\overline{n}|}^{(m)} = \frac{1-{v^{(m)}}^{mn}}{r^{(m)} \cdot v^{(m)}}$ and $\ddot{a}_{\overline{\infty}|}^{(m)} = \frac{1}{r^{(m)} \cdot v^{(m)}}$, for $r^{(m)} > 0$;
(ii) : $a_{\overline{n}|}^{(m)} = \frac{1-{v^{(m)}}^{mn}}{r^{(m)}}$ and $a_{\overline{\infty}|}^{(m)} = \frac{1}{r^{(m)}}$, for $r > 0$;
(iii): $\ddot{s}_{\overline{n}|}^{(m)} = \frac{{R^{(m)}}^{nm}-1}{r^{(m)} \cdot v^{(m)}}$;
(iv): $s_{\overline{n}|}^{(m)} = \frac{{R^{(m)}}^{mn}-1}{r^{(m)}}$.

To prove the first part of (i), we proceed exactly as we did earlier with annual payment schedule:

$$\ddot{a}^{(m)}_{\overline{n}|} = \frac{1}{m}\sum_{i=0}^{mn-1}(v^{(m)})^i = \frac{1}{m}\frac{1-v^{(m)^{mn}}}{1-v^{(m)}} = \frac{1}{m}\frac{1-v^{(m)^{mn}}}{1-\frac{1}{1+\frac{r^{(m)}}{m}}} = \frac{1-v^{(m)^{mn}}}{r^{(m)}v^{(m)}}.$$

The second assertion of (i) follows again by letting $n \to \infty$. To show the validity of the remaining formulas is left as an exercise.

Sometimes it is necessary to define annuities for durations other than integer multiples of period $\frac{1}{m}$. This can be achieved nicely using the indicator function

$$I_{[0,n)}(t) = \begin{cases} 1, & t \in [0,n) \\ 0, & \text{otherwise} \end{cases}.$$

An n-year annuity due payable m times a year can be expressed using indicator functions as

$$\ddot{a}^{(m)}_{\overline{n}|} = \frac{1}{m}\sum_{k=0}^{\infty} v^{(m)^k} I_{[0,n)}\left(\frac{k}{m}\right).$$

Suppose now that $t > 0$ (i.e. t is not necessarily an integer). The present value of a t-year annuity due payable m times a year is

$$\ddot{a}^{(m)}_{\overline{t}|} = \frac{1}{m}\sum_{k=0}^{\infty} v^{(m)^k} I_{[0,t)}\left(\frac{k}{m}\right),$$

and the present value of a t-year immediate annuity payable m times a year is

$$a^{(m)}_{\overline{t}|} = \frac{1}{m}\sum_{k=0}^{\infty} v^{(m)^k} I_{(0,t]}\left(\frac{k}{m}\right).$$

An annuity with continuous payments is obtained by letting the number of subperiods $m \to \infty$. In this case there is no difference between the cash flows of annuity due and of immediate annuity. We define a continuous cash flow as a generalization of the differential equation for provision in Definition 1.2.2. In order to do this, define the continuous time discount factor

$$v(u,t) := \exp\left(-\int_u^t \delta(v)dv\right) \tag{30}$$

and the continuous time accumulation factor

$$R(u,t) := \exp\left(\int_u^t \delta(v)dv\right), \tag{31}$$

where the force of interest δ is a piecewise continuous function.

Definition 1.4.3: A *continuous cash flow* paid into a continuous and piecewise differentiable provision V is a piecewise continuous mapping b defined on some interval $[0,T]$ such that the differential equation

$$V'(t) = \delta(t)V(t) + b(t) \tag{32}$$

is satisfied piecewise, at every point of continuity of functions δ and b on $[0,T]$.

The present value of a t-year continuous cash flow at time t_0 is

$$\int_{t_0}^{t} b(u)v(t_0,u)du. \tag{33}$$

A *continuous t-year unit annuity* is a t-year continuous constant unit cash flow (i.e. $b(u) \equiv 1$). Its present value is

$$\overline{a}_{\overline{t}|}(\delta) := \int_0^t v(0,u)du. \tag{34}$$

and its accrued provision at time t is

$$\overline{s}_{\overline{t}|}(\delta) := \int_0^t R(u,t)du. \tag{35}$$

For a constant force of interest δ

$$\overline{a}_{\overline{t}|} = \int_0^t e^{-\delta u}du = \frac{1-e^{-\delta t}}{\delta} = \frac{1-v^t}{\delta} \tag{36}$$

and

$$\overline{s}_{\overline{t}|} = \int_0^t e^{\delta(t-u)}du = e^{\delta t}\frac{1}{\delta}\left(1-e^{\delta t}\right) = \frac{R^t-1}{\delta}. \tag{37}$$

For a continuous cash flow, the provision at time $t \in [0,T]$ is

$$V(t) = R(0,t)\left(V(0) + \int_0^t b(u)v(0,u)du\right). \tag{38}$$

To see this, observe that the right side of equation (38) is the solution of the first order differential equation (32), when using the notations given in (30) and (31). Furthermore, for a continuous unit annuity with initial provision $V(0) = 0$, we have

$$V(t) = R(0,t)\overline{a}_{\overline{t}|}(\delta), \tag{39}$$

i.e. provision is the prolongated present value of the annuity, and conversely,

$$\overline{a}_{\overline{t}|}(\delta) = v(0,t)\overline{s}_{\overline{t}|}(\delta), \tag{40}$$

i.e. the present value of the annuity is the discounted value of the provision. To see this, observe that for $V(0) = 0$ and $b(u) \equiv 1$, the equation (38) takes the form

$$V(t) = R(0,t)\left(\int_0^t v(0,u)du\right) = R(0,t)\overline{a}_{\overline{t}|}(\delta), \tag{41}$$

while
(42)
$$\overline{a}_{\overline{t}|}(\delta)=\int_0^t v(0,u)du=v(0,t)\int_0^t R(0,t)v(0,u)du=v(0,t)\int_0^t R(u,t)du=v(0,t)\overline{s}_{\overline{t}|}(\delta).$$

Payments of an annuity are often linked to some index (e.g. a consumer price index). Assuming the index value increases periodically by $100\cdot k$ %, the corresponding accumulation factor is $R_k=1+k$ and the jth payment $B_j=R_k^jB$, where B is the first payment. Accrued provision for such a *geometrically increasing annuity* is

$$\sum_{j=0}^{n-1}R^{n-j}B_j=\sum_{j=0}^{n-1}R^{n-j}R_k^jB=B\cdot R\frac{R^n-R_k^n}{R-R_k}=B\cdot R\frac{R^n-R_k^n}{r-k},$$

provided that $r\neq k$ and the present value is

$$\sum_{j=0}^{n-1}v^jB_j=\sum_{j=0}^{n-1}v^jR_k^jB=B\frac{v^nR_k^n-1}{vR_k-1}=BR^{1-n}\frac{R^n-R_k^n}{r-k}.$$

Denote $\kappa=\ln(1+k)$. For a *continuous geometrically increasing annuity* with a constant force of interest $\delta\neq\kappa$ the accrued provision is

$$B\int_0^t e^{\delta(t-u)}e^{\kappa u}du=Be^{\delta t}\frac{e^{(\kappa-\delta)t}-1}{\kappa-\delta}=B\frac{e^{\kappa t}-e^{\delta t}}{\kappa-\delta}$$

and the present value is

$$B\int_0^t e^{-\delta u}e^{\kappa u}du=B\frac{e^{(\kappa-\delta)t}-1}{\kappa-\delta}=Be^{-\delta t}\frac{e^{\kappa t}-e^{\delta t}}{\kappa-\delta}.$$

For $\delta=\kappa$, the corresponding values are equal to $t\cdot e^{\delta t}\cdot B$ and $t\cdot B$, respectively.

1.7. **Internal Rate of Return.** Assume that we know the cash flow $B(t_i)$, where the payment times $t_1<t_2<\dots$ are not necessarily evenly spaced, and also we know either $S(t)$, the provision accrued up to time $t>t_1$, or $A(t)$, the present value of payments due up to time t. The *internal rate of return*[22] for the considered cash flow is the constant annual interest rate r_{IRR} satisfying either

$$S(t)=\sum_{t_j<t}B(t_j)(1+r_{IRR})^{t-t_j} \tag{43}$$

or

$$A(t)=\sum_{t_j<t}B(t_j)(1+r_{IRR})^{-t_j}. \tag{44}$$

In general (and usually in practice), r_{IRR} cannot be solved algebraically from previous equations: since time is usually measured in discrete units (day, month, year), these equations are real polynomials, possibly of high degree. The number of solutions (which may be complex numbers) is then equal to the degree of the polynomial. Numerical methods are

[22] internal rate of return = internränta = sisäinen korko(kanta)

hence usually needed if there are more than four[23] payment times or if the payment times are unequally spaced. We are interested in real-valued solutions r_{IRR} such that $r_{IRR} > -1$. Since $S(t) = R^t A(t)$, it suffices to consider equation (43).

Example: Consider the following provision: at time 0, a sum of 100 euros is loaned from the provision, at time 1 a sum of 230 euros is paid to the provision and at maturity, time 2, the provision of 132 euros is paid out. In this case the equation for accrued provision in terms of accumulation factor $R = 1 + r$ is

$$-100 \cdot R^2 + 230 \cdot R - 132 = 0.$$

This second degree polynomial has 2 roots, $R = \{1.1, 1.2\}$. i.e. we have two solutions $r_{IRR} = 10\%$ or $r_{IRR} = 20\%$.

Several real solutions larger than -1 may appear if there are alternating positive and negative cash flows. However, the following result tells us when a unique solution exists.

Proposition: Suppose $t_1 < \ldots < t_n \le t$. Equation (43) (and consequently also equation (44)) has a solution $R = 1 + r_{IRR}$ with $r_{IRR} > -1$, if $S(t) \ge 0$ and $B(t_1) > 0$. Furthermore, this solution is unique, if the provision is positive after each payment, that is,

$$S(t) = \sum_{t_j < t_k} B(t_j)(1 + r_{IRR})^{t_k - t_j} > 0, \text{ for each } k = 1, \ldots, n. \tag{45}$$

Proof: Denote

$$V(t, r) := \sum_{t_j < t} B(t_j)(1 + r)^{t - t_j} \tag{46}$$

and consider solving $V(t, r) = S(t)$ for r. Observe that V is a continuous function of r, and that $V(t, -1) = 0$ and $\lim_{r \to \infty} V(t, r) = \infty$, since by assumptions made $t > t_1$ and $B(t_1) > 0$. Hence $S(t) \ge 0$ implies that a solution exists.

To prove uniqueness under the additional assumption (45), suppose that there would exist two distinct solutions $r' > r > -1$. Then the following chain of inequalities holds:

$$\begin{aligned}
\textstyle\sum_{j=1}^n B(t_j)(1+r)^{t-t_j} &= (1+r)^{t-t_n} \left(\textstyle\sum_{j=1}^{n-1} B(t_j)(1+r)^{t_n - t_j} + B(t_n)\right) \\
&< (1+r')^{t-t_n} \left(\textstyle\sum_{j=1}^{n-1} B(t_j)(1+r)^{t_n - t_j} + B(t_n)\right) \\
&< (1+r')^{t-t_n} \left((1+r')^{t_n - t_{n-1}} \left(\textstyle\sum_{j=1}^{n-2} B(t_j)(1+r)^{t_{n-1} - t_j} + B(t_{n-1})\right) + B(t_n)\right) \\
&< \textstyle\sum_{j=1}^n B(t_j)(1+r')^{t-t_j} = S(t),
\end{aligned}$$

but then either r or r' is not a solution. Hence the solution is unique. □

In practice, we often have the situation in which the first k payments are positive and the remaining $n - k$ payments are negative. If $S(t_n) = 0$, then the provision has a unique IRR: by the previous proposition, IRR exists, and by construction the provision must be positive after each payment since the final value $S(t_n) = 0$.

[23] Polynomials of degree $d > 4$ do not have general solution formulas.

Example: Consider the cash flow where for 10 years a payment of 5 units is made at year end to a provision and after this for 5 years 12 units are taken out at the end of each year. If we wish to know what interest rate the provision should earn in order for it to be just sufficient to cover all the payments, we need to calculate the IRR of the cash flow with assumption that final provision $S(15) = 0$. That is, we need to solve

$$5 \cdot \sum_{t=1}^{10} (1+r)^{15-t} - 12 \cdot \sum_{t=11}^{15} (1+r)^{15-t} = 0.$$

for r. Solving this numerically with Excel yields $r_{IRR} = 2.436$ %. So this IRR is the interest rate the provision must earn in order for it to be sufficient to finance the outgoing payments during last 5 years. This example represents in simplified form the usual situation in a life or pension insurance contract where first premiums are paid by the insured for a specified period of time and after this the insured receives pension payments for a specified period of time.

Example: Suppose that a firm has estimated that the building of a new production plant would lead to the following cash flow of annual profit/loss:

$$B = \{-100, -150, -50, -50, 0, 10, 30, 50, 90, 120, 150, 120, 100, 100, 100\},$$

where in the first years the costs of building cause total cash flow to be negative but eventually as the new plant is finished, new products can be produced and sold, and the cash flow becomes positive. In the final years the plant is becoming obsolete and maintenance costs rise; last cash flow B_{15} represents the scrap value of the plant as it is sold, and hence the "provision"at the end of investment project is zero, $S(15) = 0$. The firm can now compute the IRR of the project from equation

$$\sum_{t=1}^{15} B_t (1+r)^{15-t} = 0.$$

, which is (using Excel) $r_{IRR} = 10.519$ %.

By calculating the internal rate of return we can compare cash flows with different payment times and maturities.

1.8. **Retrospective and Prospective Provisions; Equivalence Principle.** Consider a provision $S(t)$ defined for $t \in [0, n]$ with a force of interest $\delta(t)$ and receiving a continuous cash flow $b(t)$ which can take both positive and negative values. Dynamics of this provision are described by the differential equation

$$S'(t) = \delta(t)S(t) + b(t). \tag{47}$$

If the initial provision $S(0)$ is known, the provision is

$$S(t) = R(0,t)\left(S(0) + \int_0^t b(u)v(0,u)du\right). \tag{48}$$

As the provision is here expressed in terms of the past (before time t), it is called the *retrospective provision*.

Alternatively, the final provision $S(n)$ can be solved from equation (47) in terms of provision $S(t)$ (known at time t):

$$S(n) = R(t,n)\left(S(t) + \int_t^n b(u)v(t,u)du\right). \tag{49}$$

From this, we can solve $S(t)$:

$$S(t) = v(t,n)S(n) - \int_t^n b(u)v(t,u)du. \tag{50}$$

The provision calculated in terms of future (after time t) is called the *prospective provision*. The future is known via the contract concerning the provision, which specifies the final value, the future cash flow payments and the interest to be paid.

If the force of interest δ is constant and the cash flow $b(t)$ differs from 0 only at discrete points $0 = t_1 < t_2 < \ldots < t_m$, taking values $B(t_j)$, $j = 1, \ldots, m$, then

$$S(t) = R^t\left(S(0) + \sum_{t_i \le t} v^{t_i} B(t_i)\right) = R^t S(0) + \sum_{t_i \le t} R^{t-t_i} B(t_i) \tag{51}$$

and

$$S(t) = v^{n-t} S(n) - \sum_{t_i \ge t} v^{t_i - t} B(t_i). \tag{52}$$

Put verbally, the retrospective provision equals initial provision and paid payments with accrued interest, while the prospective provision is the difference between the final provision and the present value of future payments. Retrospective and prospective provision are naturally equal, being solutions of the same differential equation.

Two cash flows are *equivalent*, if their present values are equal under a common interest rate assumption. If the present values of two cash flows, each calculated with a different interest rate, are equal at time t, then the cash flows are *equivalent at time t*. *Equivalence principle* refers to matching two cash flows in such a way that they are equivalent at some specified moment of time.

Example: Consider an annuity paying K units at each year end for n years. The contract stipulates that the applied discount rate may be changed during the contract period, in which case the amount of annual payment is adjusted in accordance with the equivalence principle. The prevailing discount rate at the beginning of the n year period is r_1. Suppose that after the first year and payment the interest rate changes to r_2. Then the new annual payment K' satisfies

$$K \sum_{t=1}^{n-1} v_1^t = K' \sum_{t=1}^{n-1} v_2^t,$$

which is equivalent to

$$K' = \frac{K \cdot a_{\overline{n-1|}}(r_1)}{a_{\overline{n-1|}}(r_2)}.$$

In general, given discount factors v_i and payments $B_i(t^i_j)$, $i = 1, 2$ and $j^i = 1, 2, \ldots, m^i$, where $i = 1$ is before the change and $i = 2$ is after the change, to satisfy the equivalence principle we must have

$$\sum_{j=1}^{m^1} v_1^{t^1_j} B_1(t^1_j) = \sum_{j=1}^{m^2} v_2^{t^2_j} B_2(t^2_j).$$

For continuous payments and forces of interest this becomes

$$\int_t^{n_1} b_1(u) v_{\delta_1}(t, u) du = \int_t^{n_2} b_2(u) v_{\delta_2}(t, u) du.$$

1.9. **Duration.** It is obvious that the present value of a cash flow depends on the applied discount rate. As interest rates on the markets are not constant in time, it is important to know how the present value of a cash flow changes when the discount rate changes. This can be measured with *duration*.

Definition 1.7.1: Let $B(t)$ be a cash flow and δ the force of interest. *Duration* of $B(t)$ is

$$\bar{D}(\delta) := \frac{\int_0^\infty t \cdot v^t B(t) dt}{\int_0^\infty v^t B(t) dt} =: \mathbb{E}(T), \tag{53}$$

where $v = e^{-\delta}$ and T is a random variable whose density function is

$$f_T(t) = \frac{v^t B(t)}{\int_0^\infty v^u B(u) du}. \tag{54}$$

If $B(t)$ is a continuous cash flow, the integral is interpreted as a Riemann integral over the lifetime of the cash flow; if $B(t)$ is a discrete cash flow, the integral is interpreted as a sum over payment times with nonzero cash flow.

Duration of a cash flow is a weighted average of the cash flow's payment times where the weights are the present values of payments. For a discrete cash flow $B(t_i)$, $i = 1, \ldots, n$, with known payment times and magnitudes of payments, duration equals

$$\bar{D}(\delta) = \frac{\sum_{j=1}^n t_j v^{t_j} B(t_j)}{\sum_{j=1}^n v^{t_j} B(t_j)}, \tag{55}$$

Observing that in this case $e^{-\delta} = e^{-\ln(1+r)} = \frac{1}{1+r}$, we can write this also as a function of the discount rate r as

$$D(r) := \bar{D}\left(\ln(1+r)\right) = \frac{\sum_{j=1}^n t_j \frac{B(t_j)}{(1+r)^{t_j}}}{\sum_{j=1}^n \frac{B(t_j)}{(1+r)^{t_j}}}. \tag{56}$$

Denote now by $P(B, \delta)$ the present value of cash flow B with a force of interest δ. The (infinitesimal) relative change in the present value of the cash flow due to a (infinitesimal) change in the force of interest δ is

$$\frac{1}{P(B,\delta)}\frac{\partial P(B,\delta)}{\partial \delta}. \tag{57}$$

We have the following result:

$$\bar{D}(\delta) = -\frac{1}{P(B,\delta)}\frac{\partial P(B,\delta)}{\partial \delta}. \tag{58}$$

This follows easily by observing that

$$\frac{\partial P(B,\delta)}{\partial \delta} = \frac{\partial}{\partial \delta}\left[\int_0^\infty e^{-\delta t} B(t)dt\right] = -\int_0^\infty t \cdot e^{-\delta t} B(t)dt,$$

which combined with the definition of duration in (53) implies (58). (Provided, of course, that interchanging the order of integration and differentiation is allowed. In practical applications this is nearly always the case.) Hence for a small change $\Delta\delta$ in the force of interest δ, we have an approximation for the corresponding relative change in the present value of cash flow B:

$$\frac{\partial P(B,\delta)}{P(B,\delta)} = -\bar{D}(\delta)\partial\delta \Rightarrow \frac{\Delta P(B,\delta)}{P(B,\delta)} \approx -\bar{D}(\delta)\Delta\delta. \tag{59}$$

This approximation is a widely used application of the duration concept.

We can also consider the sensitivity of duration $\bar{D}(\delta)$ to changes in the force of interest δ:

$$\begin{aligned}
\frac{\partial \bar{D}(\delta)}{\partial \delta} &= \frac{\partial}{\partial \delta}\left[\frac{\int_0^\infty t v^t B(t)dt}{\int_0^\infty v^t B(t)dt}\right] \\
&= \frac{1}{\int_0^\infty v^t B(t)dt} \cdot \frac{\partial}{\partial \delta}\left[\int_0^\infty t v^t B(t)dt\right] - \frac{\int_0^\infty t v^t B(t)dt}{\left(\int_0^\infty v^t B(t)dt\right)^2} \cdot \frac{\partial}{\partial \delta}\left[\int_0^\infty v^t B(t)dt\right] \\
&= \frac{-\int_0^\infty t^2 v^t B(t)dt}{\int_0^\infty v^t B(t)dt} + \frac{\left(\int_0^\infty t v^t B(t)dt\right)^2}{\left(\int_0^\infty v^t B(t)dt\right)^2} = \bar{D}(\delta)^2 - \mathbb{E}(T^2).
\end{aligned}$$

So we see that in terms of the random variable T defined in Definition 1.7.1, we have

$$\frac{\partial \bar{D}(\delta)}{\partial \delta} = \bar{D}(\delta)^2 - \mathbb{E}(T^2) = -Var(T). \tag{60}$$

We can also consider duration as a function of the interest rate r according to equation (56). In this case the relative change of the value of cash flow B due to a change in the interest rate r is

$$\frac{1}{P(B,r)}\frac{\partial P(B,r)}{\partial r} = \frac{\partial \delta}{\partial r}\frac{1}{P(B,\delta)}\frac{\partial P(B,\delta)}{\partial \delta} = -\frac{1}{1+r}\bar{D}(\delta) = -\frac{1}{1+r}D(r), \tag{61}$$

since $\frac{\partial \delta}{\partial r} = \frac{\partial}{\partial r}\ln(1+r) = \frac{1}{1+r}$.

Example: Consider the cash flow $B(t) = e^{0.05t}$, $t \in [0,4]$ when the force of interest $\delta = 0.03$. The present value of this cash flow is

$$P(B, 0.03) = \int_0^4 e^{-0.03t} e^{0.05t} dt = \frac{e^{0.02\cdot 4} - e^{0.02\cdot 0}}{0.02} = 4.164,$$

and its duration is

$$\bar{D}(0.03) = \frac{\int_0^4 te^{-0.03t} e^{0.05t} dt}{P(B, 0.03)} = 2.0267.$$

If the force of interest δ goes up with 1 percentage point to 0.04, the relative change in the present value of cash flow B is approximately

$$-\bar{D}(0.03) \cdot \Delta\delta = -2.0267 \cdot 0.01 = -0.020267,$$

that is, the present value decreases by approximately 2.0 %.

Example: Consider the discrete cash flow $B = \{0.1, 0.1, 0.1, 1.1\}$ where payments occur at the ends of years 1, 2, 3 and 4, when the interest rate $r = 0.03$. The present value of this cash flow is

$$P(B, 0.03) = \sum_{t=1}^{3} \frac{0.1}{1.03^t} + \frac{1.1}{1.03^4} = 1.2602,$$

and its duration is

$$D(0.03) = \frac{\sum_{t=1}^{3} \frac{0.1t}{1.03^t} + \frac{1.1\cdot 4}{1.03^4}}{P(B, 0.03)} = 3.5467.$$

If the interest r goes down with 1 percentage point to 0.02, the relative change in the present value of cash flow B is approximately

$$-\frac{1}{1+r} \cdot D(0.03) \cdot \Delta r = -\frac{1}{1.03} \cdot 3.5467 \cdot (-0.01) = 0.034434,$$

that is, the present value increases by approximately 3.4 %.

1.10. **Some Financial Instruments and Investment Opportunities.** In modern financial markets, a bewildering array of financial instruments exists, from simple conventional fixed coupon rate bonds to extremely complex asset-backed securities and financial derivatives. In this subsection we look at the valuation of some of the most common financial instruments and their risk and return characteristics. More specifically, we will consider

- **a:** fixed income instruments, further subdivided into loans, money market instruments and bonds,
- **b:** stocks (or equities),
- **c:** real estate investments,
- **d:** some alternative investment opportunities such as hedge funds, commodities and private equity,
- **e:** some of the most common financial derivatives.

We begin with *fixed income instruments*, which provide a stream of fixed payments to the holder of the instrument. Fixed income instruments can be broadly divided into loans, money market instruments and bonds. Observe that term "fixed income" refers to the fact that the payment stream is contractually fixed in advance – but it does not mean risk-free, as the issuer may be subject to default risk. Even in absence of default risk, the return from a fixed income instrument is guaranteed only if the instrument is held to maturity: the market prices of these instruments fluctuate with interest rates (and with other factors) in the secondary markets.

1.10.1. *Loans.* As was previously observed, a *loan*[24] is an investment for the lender. The lender gives a part of his/her assets to the borrower temporarily, in return for interest payments received from the borrower during the lifetime of the loan. The borrower pays back the loan during the lifetime of the loan. There are several options as to how this repayment is scheduled. Repayments of a loan are called *amortizations*[25]. We will next consider the most common amortization schedules.

In a *loan with a fixed amortization schedule* all amortization payments by the borrower are equal. In addition, the borrower pays at each amortization payment date an interest payment equal to accrued interest for the remaining principal. To illustrate, suppose that amortization payments of a loan of amount L are paid m times a year for n years. Then the constant amortization payment is $T = L/(mn)$, and the interest payment at i'th payment date is $\frac{r^{(m)}}{m} \cdot (L - i \cdot T)$, so that i'th payment equals

$$S_i = T + \frac{r^{(m)}}{m} \cdot (L - (i-1) \cdot T), \text{ for } i = 1, \ldots, mn.$$

If discounted with the loan's interest rate, the present value of the cash flow of the loan, $\{S_k\}$, $k = 1, \ldots, mn$, is equal to loan principal L.

In an *annuity loan*[26], the sum of amortization and interest payment is constant. Denoting this sum by A, we have by equivalence principle that for a n-year loan with annual payments

$$L = \sum_{k=1}^{n} A \cdot v^k = A \frac{1 - v^n}{r} = A \cdot a_{\overline{n|}}. \tag{62}$$

We can solve this for annual payment A to obtain

$$A = \frac{L}{a_{\overline{n|}}} = L \cdot \frac{r}{1 - v^n}. \tag{63}$$

The value of an annuity loan between kth and $(k+1)$th payments is

$$V(k+t) = A \cdot a_{\overline{n-k|}} \cdot R^t = A \cdot s_{\overline{n-k|}} v^{n-k} R^t = A \cdot s_{\overline{n-k|}} \cdot v^{n-(k+t)},$$

[24] loan = lån = laina
[25] amortization = amortering = kuoletus, lyhennys
[26] annuity loan = annuitetslån = annuiteettilaina

where $0 \leq t < 1$. Interpretation of the first equality is that loan value is the present value of future payments; interpretation of the last equality is that loan value is the accrued provision from time $k+t$ up to maturity n, discounted to time $k+t$.

For an annuity loan, a change in the applied interest rate implies that either the payment A or maturity n must change. Denoting the values before the change by x_1 and values after the change by x_2 for $x \in \{r, n, A\}$, for the loan value to remain unchanged we must have

$$A_1 \cdot a_{\overline{n_1-k}|}(r_1)R_1^t = A_2 \cdot a_{\overline{n_2-k}|}(r_2)R_2^t,$$

where $a_{\overline{n_1-k}|}(r)$ means that the annuity is calculated with interest rate r. From this equivalence the desired quantity can be solved by fixing all others. Usually holding A constant is not possible: the last payment A_{n_2} will be smaller than other payments and can be solved from

$$A \cdot a_{\overline{n_1-k}|}(r_1)R_1^t = A \cdot a_{\overline{n_2-1-k}|}(r_2)R_2^t + A_{n_2}v_2^{n_2-(k+t)}.$$

Annuity payment A can be decomposed into interest I_k and amortization L_k as follows:

$$A = I_k + L_k = r \cdot A \cdot a_{\overline{n-k+1}|} + A \cdot v^{n-k+1}. \tag{64}$$

In a *bullet loan*[27] the principal is repaid at maturity and interest payments are made m times a year. Principal L is related to nominal interest rate $r^{(m)}$ paid m times a year for n years by equation

$$L = \sum_{k=1}^{mn} \frac{r^{(m)}}{m} L \cdot {v^{(m)}}^k + L \cdot {v^{(m)}}^{mn} = L \cdot \left(r^{(m)} a_{\overline{n}|}^{(m)} + {v^{(m)}}^{nm} \right). \tag{65}$$

1.10.2. *Money Market Instruments. Money market*[28] consists of financial instruments with maturity less than or equal to one year. These instruments are issued by governments and corporations to satisfy their need of short-term funding. Money market instruments do not make any intermediate interest payments: the buyer (lender) buys the instrument at market price, thus borrowing an amount equal to the price to the issuer (borrower), and at maturity the buyer receives the face value (which is greater than the price) of the instrument from the issuer. The interest earned is hence equal to the difference between the price and the face value. Money market interest rates are simple interest rates, as no interest on interest is earned. Securities have standardized maturities with most important ones being 1-, 2-, 3-, 6-, 9- and 12-month maturities.

The price of a money market security with maturity T and cash flow at maturity equal to C_T is

$$P_{MM}(0,T) = \frac{C_T}{1 + r_s T}, \tag{66}$$

where r_s is the simple annual interest rate for the period in question.

[27]bullet loan = bulletlån = bullet-laina
[28]money market = penningsmarknad = rahamarkkinat

Example: Interest rate for a certificate of deposit (CD) with 30 days to maturity and face value of 1000000 euros is 3.5 %. The price of the CD is

$$\frac{1000000}{1 + 0.035 \cdot \frac{30}{360}} = 997091.82.$$

A significant difference between money market instruments and loans is that money market instruments have very liquid secondary markets, where the holders of instruments can sell them to other investors on the market before maturity. Usual loans are usually not easily transferrable but must be held until maturity. Naturally, while the cash flows from a loan are fixed and there is no capital loss unless the borrower defaults, selling a money market instrument before maturity at the market can generate both capital losses or gains, depending on the level of the market price.

1.10.3. *Bonds.* On *bond*[29] *markets*, governments and corporations borrow funds for long term; maturities range from one year to several decades (there exist even so-called perpetual bonds, whose principal is never repaid but coupon payments are made to lenders indefinitely).

Here is how the process works: the entity in need of capital issues securities, *bonds*, with a specified *face value*[30], which any investor can buy. In return for loaning thus the price of the bond to the issuer, the buyer usually receives coupon payments periodically (annually or semiannually, say) during the time to maturity of the bond. Coupon payments are usually specified as a percentage of the face value (coupon rate). At maturity date, the issuer pays to the buyer the face value and the last coupon. Similar to money market instruments and unlike loans, bonds have liquid secondary markets: the buyer of a bond does not need to hold the bond until maturity but instead can sell it on the market for its market price, which operation may generate capital gains or losses. However, market size, depth and liquidity varies: usually developed governments' bonds have quite large and liquid markets while in some countries the corporate borrowing is mainly done via banks and the corporate bond market is small and undeveloped.

We will here consider the pricing of option-free bonds, which constitute the majority of the market. Many bonds, however, have embedded options. Some examples are convertible bonds, callable bonds and putable bonds. Pricing bonds with embedded options correctly requires pricing the embedded options, i.e. option pricing methods are needed for that.

Pricing of bonds begins from the basic case of a *zero-coupon bond*: a bond which pays its face value of 1 euro at maturity and has no coupon payments during the time to maturity. A bond can be decomposed into a sum of zero-coupon instruments, and each of this instruments can be valued by discounting with the appropriate spot rate. The present

[29] bond = masskuldebrevslån = joukkovelkakirjalaina

[30] Face value =nominalvärde = nimellisarvo

value at time t of a bond with maturity T and time to maturity $T-t$ is hence

$$P(t,T)=\sum_{k=0}^{n-1}CF_{\tau+\frac{k}{m}}e^{-(\tau+\frac{k}{m})s(t,\tau+\frac{k}{m})}=\sum_{k=0}^{n-1}N\cdot\frac{r_{cpn}^{(m)}}{m}e^{-(\tau+\frac{k}{m})s(t,\tau+\frac{k}{m})}+Ne^{-(\tau+\frac{n-1}{m})s(t,\tau+\frac{n-1}{m})}, \tag{67}$$

where N is the face value, $s(v,u)$ is the continuous spot rate at time v with maturity u, $r_{cpn}^{(m)}$ is the fixed annual simple coupon interest rate and τ is the time in years to the next coupon payment date from time t. Time unit $1/m$ is the time in years between the periodic payments. To simplify, we set $t=0$ and consider discrete spot rates:

$$P(0,T)=\sum_{k=0}^{n-1}\frac{CF_{\tau+\frac{k}{m}}}{(1+s_{\tau+\frac{k}{m}})^{\tau+\frac{k}{m}}}=\sum_{k=0}^{n-1}\frac{N\cdot\frac{r_{cpn}^{(m)}}{m}}{(1+s_{\tau+\frac{k}{m}})^{\tau+\frac{k}{m}}}+\frac{N}{(1+s_{\tau+\frac{n-1}{m}})^{\tau+\frac{n-1}{m}}}, \tag{68}$$

where s_w is now the current spot rate with maturity w.

There are also bonds with a floating coupon rate. Suppose that the number of coupon payments until the adjustment date is h. The floating rate is then adjusted immediately after h:th coupon payment in such a way that the price of the bond equals its face value, that is, if v is the time after hth coupon payment, then $P_{floating}(v,T)=N$. Since the price $P_{floating}(v,T)$ is also the present value of the bond's remaining cash flows at time v, the present value of a bond paying floating rate coupons is

$$\begin{aligned}
P_{floating}(0,T) &= \textstyle\sum_{k=0}^{n-1}\frac{N\cdot\frac{r_{floating}^{(m)}}{m}}{(1+s_{\tau+\frac{k}{m}})^{\tau+\frac{k}{m}}}+\frac{N}{(1+s_{\tau+\frac{n-1}{m}})^{\tau+\frac{n-1}{m}}}\\
&= \textstyle\sum_{k=0}^{h-1}\frac{N\cdot\frac{r_{floating}^{(m)}}{m}}{(1+s_{\tau+\frac{k}{m}})^{\tau+\frac{k}{m}}}+\sum_{k=h}^{n-1}\frac{N\cdot\frac{r_{floating}^{(m)}}{m}}{(1+s_{\tau+\frac{k}{m}})^{\tau+\frac{k}{m}}}+\frac{N}{(1+s_{\tau+\frac{n-1}{m}})^{\tau+\frac{n-1}{m}}}\\
&= \textstyle\sum_{k=0}^{h-1}\frac{N\cdot\frac{r_{floating}^{(m)}}{m}}{(1+s_{\tau+\frac{k}{m}})^{\tau+\frac{k}{m}}}+\frac{P_{floating}(\tau+\frac{h-1}{m},T)}{(1+s_{\tau+\frac{h-1}{m}})^{\tau+\frac{h-1}{m}}}\\
&= \textstyle\sum_{k=0}^{h-1}\frac{N\cdot\frac{r_{floating}^{(m)}}{m}}{(1+s_{\tau+\frac{k}{m}})^{\tau+\frac{k}{m}}}+\frac{N}{(1+s_{\tau+\frac{h-1}{m}})^{\tau+\frac{h-1}{m}}}.
\end{aligned} \tag{69}$$

It is immediately clear that knowing the price of a n-year zero coupon bond is equivalent to knowing the n-year spot rate; this means that the term structure of interest rates can be derived from bond prices.

As stated, when calculating the present value of a bond, the cash flows are discounted with different interest rates depending on the maturity of the individual payment. The constant compounded discount rate equating the price of the bond with the discounted value of its cash flows is called the *yield*[31] (or *yield to maturity*) of the bond. Yield y of a bond is the solution to equation

$$P(0,T) = \sum_{k=0}^{n-1} \frac{CF_{\tau+\frac{k}{m}}}{(1+s_{\tau+\frac{k}{m}})^{\tau+\frac{k}{m}}} = \sum_{k=0}^{n-1} \frac{CF_{\tau+\frac{k}{m}}}{(1+y)^{\tau+\frac{k}{m}}}. \tag{70}$$

Payment schedule (cash flows and their timing) is specified in bond contract, and the bond price is observed in the market. The yield is also called the bond's internal rate of return. It is a convenient summary measure of the interest rate level of a bond.

The yield is not the same as the total return from a bond investment. The total return comes from coupon payments, income from reinvestment of the coupons and capital gain or loss when the bond is held to maturity or sold. Yield and total return are equal only if the coupon payments can be reinvested at the yield rate and the bond is held to maturity.

It is clear that the bond price depends on the interest rate(s). Hence changes in interest rates lead to changes in the price of the bond, and if the bond is sold before maturity, the investor may realize a loss or a gain. In other words, the cash flow of a bond is subject to *interest rate risk*. In general, bond prices and bond yields move in opposite directions: when yields rise, prices fall and vice versa.

As we saw in Section 1.9, the sensitivity of bond price (present value of a cash flow) to small changes in the yield (constant discount rate) can be measured by *duration*.

The average cash flow maturity of a bond is given by the *Macaulay duration*

$$D_{Mac} := \frac{1}{P_{bond}} \sum_{k=0}^{n-1} \frac{(\tau+\frac{k}{m})CF_{\tau+\frac{k}{m}}}{(1+y)^{\tau+\frac{k}{m}}}.$$

It is a trivial exercise to show that this is equal to maturity for a zero-coupon bond and for a coupon bond always less than maturity.

To determine the approximate change in price for a small change in yield, so-called *modified duration* is used. It is defined as

$$D_{Mod} := \frac{1}{1+y} D_{Mac},$$

and a comparison with Section 1.9 quickly shows that the change in bond price due to an infinitesimal change in yield dy is

$$dP_{bond} = -D_{Mod} \cdot P_{bond} \cdot dy.$$

For small changes in yield, approximation with modified duration is a good measure of interest rate sensitivity. However, for yield increases duration overestimates the price decrease and for yield decreases it underestimates the price increase. For larger yield changes the approximation is not very good.

The approximation can be refined by considering the Taylor expansion of the bond price:

$$dP_{bond} = \frac{dP_{bond}}{dy} dy + \frac{1}{2} \frac{dP_{bond}}{dy^2} (dy)^2 + R_3.$$

[31] yield = avkastning = tuotto

The first term in the expansion is the price change due to duration: essentially the approximation using duration is based on truncating the Taylor series expansion after the first order term. The second term is called the price change due to *convexity.* Convexity is the derivative of duration with respect to yield, and it measures the curvature of the bond price. It can be defined as

$$Convexity = \frac{dP_{bond}}{dy^2}\frac{1}{P_{bond}} = \frac{1}{P_{bond}}\frac{1}{(1+y)^2}\sum_{k=0}^{n-1}\frac{(\tau+\frac{k}{m})(\tau+\frac{k}{m}+1)CF_{\tau+\frac{k}{m}}}{(1+y)^{\tau+\frac{k}{m}}}.$$

Price change due to convexity is then

$$dP_{bond} = \frac{1}{2}\frac{dP_{bond}}{dy^2}(dy)^2 = \frac{1}{2}\cdot P_{bond}\cdot Convexity\cdot(dy)^2.$$

Adding together the price changes due to duration and convexity, we obtain a second order approximation for the price change due to change in yield:

$$dP_{bond} = -D_{Mod}\cdot P_{bond}\cdot dy + \frac{1}{2}\cdot P_{bond}\cdot Convexity\cdot(dy)^2.$$

As convexity measures the change in duration when yield changes, a high convexity can be interpreted as larger dispersion of a cash flow.

Duration and convexity only measure the sensitivity of the bond price for parallel shifts of the yield curve (i.e. term structure). However, the term structure may (and does) also change in such a fashion that the percentage change is not equal for all maturities. In this case two bonds with identical durations may not behave similarly when there is a change in the term structure.

Previous valuation analysis assumes the cash flows from a bond to be deterministic, and as such it applies only to *default-free bonds.* While payments from a bond are fixed contractually, they are subject to default risk: bond issuer's default may lead to loss of coupon payments or principal. Hence the creditworthiness of the issuer is an important factor in pricing bonds.

Government bonds in broad sense include, in addition to debt issued by national governments (*sovereign debt*), also debt issued by local governments (such as, for example, State of California in the U.S.A., or state of Bavaria in Germany), municipalities and certain public institutions. Essential common feature is that the borrower is a *public* institution. Government bonds of national governments of developed nations have been considered (in theory) essentially default-free, i.e. the chance of the borrower defaulting on debt is thought to be negligible or at least very low. This is based usually on the argument that national governments have either the power to levy taxes or to print money (member states of the E.U. do not have the second option), and hence they can always finance their debt – if they wish to do so. There are fairly recent cases of defaults on sovereign debt, especially among developing countries (for example, Russia in 1998, Argentina in 2002); several of these are related more to a lack of willingness than of ability to pay. However, for valuation purposes traditionally government bonds are priced as if they were default-free, i.e. the risk-free interest rate for the bond's maturity is used in calculating the present value of the bond.

Default risk is explicitly taken into account in the pricing of *corporate bonds*, i.e. bonds issued by private corporations. This is implemented by using a risk-adjusted (i.e. higher than the risk-free rate) interest rate in present value calculations. Higher discount rate decreases the price of the bond and increases its yield, thus making it more attractive to investors, who require a higher yield from corporate debt to compensate for its higher risk of default in comparison with government debt. The market's assessment of the severity of default risk is characterized by the *credit spread* of the bond – that is, how much higher is the interest rate of the bond than the risk-free rate of same maturity. Spreads vary considerably between different issuers: while a large, financially sound international corporation with a high credit rating may have a spread of 0.1 % (say), a small corporation on the brink of bankruptcy with a low credit rating might have a spread in excess of 20 %, say. In addition, credit spreads vary in time depending on the risk sentiment on the markets and the general economic outlook: during the recent financial turmoil in 2008-09, even the spreads of relatively sound issuers widened considerably as markets panicked and just about everything except maybe U.S. Treasury debt was considered unsafe. This phenomenon of investors pulling their money out of risky investments and investing into so-called *safe haven* instruments (e.g. developed market countries' sovereign debt, gold) during periods of financial distress is known as *flight to quality*.

Despite the previous distinction between government and corporate debt, credit rating agencies rate both governments and corporations, and spreads for different countries may differ considerably (compare Greece and Germany at the moment).

Because of credit risk, holders of corporate debt may require as a security fixed assets, real property or other securities. In case this security is fixed assets or real property, the bond is called a *mortgage bond*. The bondholder has a legal right to sell mortgaged property to satisfy unpaid obligations to bondholders.

In a *collateral bond*, the issuer of the bond pledges financial assets such as stocks or other bonds as collateral to satisfy the bondholders' need for security.

Debenture bonds are not secured by specific pledged assets, but have a claim on all non-pledged assets of the issuers and on pledged assets to the extent that the value of the pledged assets is greater than what is needed to satisfy secured creditors. *Subordinated debenture bonds* rank after secured debt, debenture bonds and some other general creditors in their claim on the issuer's assets and earnings. The type of security influences the cost to the issuer, i.e. the safer the security is considered, the lesser the excess yield (risk premium above the risk-free rate) investors require and hence the lesser the cost of financing to the issuer.

A *guaranteed bond* is a bond guaranteed by another entity. Safety of such a bond naturally depends on the financial capability of the guarantor.

Bonds with embedded options (to the bondholder) include *convertible bonds*, which can be converted to a predetermined number of shares of the common stock of the bond issuer, *exchangeable bonds*, which can be exchanged for shares of some other firm than the issuer, bonds with *warrants* attached (a right to purchase a designated security at a specified price), and *putable bonds*, which can be sold back to the issuer at a designated price and maturity. In a *callable bond*, the issuer has the right to retire ("call") all or part of the issue before maturity date.

1.10.4. *Stocks.* In *stock markets*[32], corporations raise long-term capital by issuing shares. Unlike debt instruments, shares do not have a fixed maturity. A share of stock (or equity) represents a residual claim on the issuer's future earnings. In comparison with bonds, stocks have a higher risk but also a higher potential return. In case of bankruptcy of the issuing corporation, bondholders and other creditors are first paid and the residual assets (if any) are divided among the shareholders: hence the shareholders may easily lose their whole investment if the issuer goes bankrupt. On the other hand, no matter how well the issuer does, bondholders never receive more than the agreed interest payments and the repayment of principal. Shareholders, in contrast, have (in principle) unlimited upside potential to participate in the issuer's success: if the share price appreciates tenfold, the holder may, by selling his shares, realize a 900 % return. In addition to appreciation of shares' value, the shareholders usually receive annual dividends from the issuer: the size of the dividend payment may vary.

As common stock represents an ownership position in a company, the shareholders have voting power proportionate to the number of shares in shareholder meetings, which among other things elects the board of directors.

Important characteristic of stocks is *limited liability*: even though shareholders own the company, their liability is restricted to the value of their shares. Shareholders cannot be forced to provide funds for settling of the company's obligations in case of a bankruptcy.

The sum of the cumulative retained earnings and other stockholder's equity in a company's balance sheet is called the book value of equity. The book value per share is obtained by dividing the book value of equity by the number of shares outstanding.

The market value of a stock is the price at which buyers and sellers trade the stocks in an open marketplace.

Pricing of stocks is more difficult than pricing of bonds, as the cash flows are not contractually fixed and the return on investment is generated by unknown future dividend payments and unknown future changes in the stock price (appreciation or depreciation). Theoretically, if the stock is not sold, its value should be equal to the (expected) present value of all future dividends; that is, given a future (infinite) dividend flow $\{D_k\}_{k=1}^{\infty}$ and a discount rate r, the value of the stock is

$$V = \sum_{k=1}^{\infty} v^k D_k.$$

As this method considers value from the stockholder's perspective, the applied discount rate should reflect the shareholder's opportunity cost of providing equity capital to the firm; we may denote such a discount rate by r_{stock}. This valuation method is called the *dividend discount model*. In the special case that $D_k = (1+\mu)D_{k-1}$, $k \geq 1$ and $D_0 = D$,

[32] stock/equity market = aktiemarknad = osakemarkkinat

where the annual growth $\mu < r$, the previous formula simplifies to

$$V = D \cdot \frac{1}{1 - \frac{1+\mu}{1+r}} = D\frac{1+r}{r-\mu}.$$

Often in this kind of valuation models the future is split into two periods: an explicit forecasting period, say T, and the time after this. For forecasting period, the cash flows are forecasted explicitly, while for the period after the forecasting period the company is assumed to have a specified *continuing value* CV_{T-1}. This leads to a finite sum in the valuation equation:

$$V = \sum_{t=0}^{T-1} \frac{CF_{\tau+t}}{(1+r_{stock})^{\tau+t}} + \frac{CV_{\tau+T-1}}{(1+r_{stock})^{\tau+T-1}}.$$

An extension of the dividend discount model is the *value added discount model*, which takes into account not only dividends but also the increase in the company's undivided assets (which belong to the shareholders as owners of the company), that is, the increase in the book value of equity. In this model the value of a stock is

$$\begin{aligned} V &= \textstyle\sum_{t=0}^{\infty} \frac{VA_{\tau+t}}{(1+r_{stock})^{\tau+t}} \\ &= BV_{\tau-1} + \textstyle\sum_{t=0}^{T-1} \frac{E_{\tau+t} - rBV_{\tau+t-1}}{(1+r_{stock})^{\tau+t}} + \frac{CV_{\tau+T-1}}{(1+r_{stock})^{\tau+T-1}} \\ &= BV_{\tau-1} + \textstyle\sum_{t=0}^{T-1} \frac{D_{\tau+t}}{(1+r_{stock})^{\tau+t-1}} + \textstyle\sum_{t=0}^{T-1} \frac{E_{\tau+t}(1-p_{\tau+t} - rBV_{\tau+t-1})}{(1+r_{stock})^{\tau+t-1}} + \frac{CV_{\tau+T-1}}{(1+r_{stock})^{\tau+T-1}}, \end{aligned}$$

where

$$BV_{\tau+t-1} = \begin{cases} BV_{\tau+t-2} + E_{\tau+t-1}(1-p_{\tau+t-1}) & , t \geq 1 \\ BV_{\tau-1} & , t = 0, \end{cases}$$

B_u is the book value of equity at time u, p_u is the payout ratio at time u and E_u is earnings on period u.

Previous valuation models consider the value from the shareholder's viewpoint. However, firms usually finance their operations both by equity and debt. From this point of view, an alternative valuation method taking into account also the debtor's (bondholder's) viewpoint is *free cash flow discount model.* Free cash flow (FCF) is the cash flow generated by the firm's operations and available to all firm's capital providers. It equals earnings before interest and taxes (EBIT), less taxes and investments in working capital and other assets. Observe that financing related cash flows such as dividends and interest expenses are not incorporated. The value of the company equals the present value of expected future free cash flows (the operating value of the company), less the present value of the company's debt and adjusted for the present value of any non-operating assets or liabilities (such as options). The applied discount rate needs to reflect the opportunity cost to all providers of capital weighted by the relative contribution to the firm's total capital. This discount rate is called the *weighted average cost of capital* (WACC) and can be defined as

$$r_{WACC} = \frac{S}{S+D} r_{stock} + \frac{D}{S+D}(1 - r_{tax}) r_{debt},$$

where S is the market value of equity, D is the market value of interest bearing debt and r_{tax} the company tax rate. In reality, the capital structure varies in time and the weighting ratios of the previous formula could be chosen to match the company's target capital structure. Free cash flow forecasts are then developed for a set of different performance scenarios thought to describe well and extensively the future possibilities, and to each of the scenarios a probability of realization is attached. Expected future cash flow FCF is then the probability-weighted average of forecasts taken over scenarios. Hence the operating value of the firm is defined as

$$V_{oper} = \sum_{t=0}^{T-1} \frac{FCF_{\tau+t}}{(1 + r_{WACC})^{\tau+t}} + \frac{CV_{\tau+T-1}}{(1 + r_{WACC})^{\tau+T-1}},$$

where FCF_u is the probability-weighted average of free cash flow forecasts for period u. The present value of the company is then

$$V = V_{oper} - D \pm NO,$$

where D is the present value of company's debt and NO is the present value of non-operating assets and liabilities.

However, mostly appreciation in value to levels above the originally paid price forms a significant part of the total return in a stock investment. The development of the stock price, in turn, is influenced by a multitude of market factors.

Dividends and future stock prices depend on both company-specific factors and on general economic factors. The uncertainty related to the future development of these factors gives rise to risks divided commonly into *non-systematic*, company-specific risk related strictly to the specific company and *systematic*, general economic risks influencing all sectors and companies. Examples of the latter risks could be interest rate level, inflation, tax reforms and political changes.

The relationship between return and risk is often measured using the *Capital Asset Pricing Model* (CAPM). It postulates that the average return of a specific stock $\bar{r}$ equals the risk free return r_f plus the stock's β multiplied by the market price of risk (or market risk premium):

$$\bar{r} = r_f + \beta(\bar{r}_M - r_f), \tag{71}$$

where $\bar{r}_M$ denotes the expected return of the overall stock market portfolio. A company's β describes how the stock is related to the market return and it is estimated from stock and market index returns. A $\beta > 1$ indicates an aggressive stock which is more volatile than the overall market, a $\beta < 1$ indicates a defensive stock which is less volatile than the overall market. Observe that CAPM equates risk with volatility or standard deviation, an assumption open to severe criticism. Writing the linear market model of CAPM in stochastic form as follows:

$$\tilde{r}_t = \alpha + \beta \cdot \tilde{r}_t^M + \tilde{\epsilon}_t,\ t = 1, 2, \ldots, n, \tag{72}$$

where $\tilde{r}_t$ is the (stochastic) return of the stock, $\tilde{r}_t^M$ is the market return and the error terms $\tilde{\epsilon}_t$ are independent and identically $N(0, \sigma^2)$-distributed random variables, we see that the total risk of the stock (measured by its variance in the market model) can be decomposed into market (systematic) risk and unique (non-systematic) risk:

$$Var(\tilde{r}_t) = \beta^2 \sigma^2_{\tilde{r}^M} + \sigma^2_{\tilde{\epsilon}}.$$

An assumption often made is that continuously compounded one-period stock returns $\tilde{\delta}_i$ are independent and identically normally distributed random variables, i.e. $\tilde{\delta}_i \sim N(\mu, \sigma^2)$. As $\tilde{\delta}_i = \ln(1 + \tilde{r}_i)$, we see that total returns $1 + \tilde{r}_i$ are lognormally distributed random variables. Hence the mean and the variance of simple returns are

$$\mathbb{E}[\tilde{r}_i] = e^{\mu + \sigma^2/2} - 1 \text{ and } Var[\tilde{r}_i] = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1).$$

For a T-year period, the mean and variance of the continuously compounded returns of the stock are μT and $\sigma^2 T$, where μ and σ^2 are the annual expected return and variance. Denote the random stock price at time T by $\tilde{S}_T$ and the current market price by S_0. Then

$$\tilde{r}_T = \frac{\tilde{S}_T - S_0}{S_0} \Rightarrow \tilde{\delta}_T = \ln(\tilde{S}_T / S_0).$$

Hence

$$\mathbb{E}[\tilde{S}_T] = S_0 e^{(\mu + \sigma^2/2)T} \text{ and } Var[\tilde{S}_T] = S_0^2 e^{2\mu T + \sigma^2 T}(e^{\sigma^2 T} - 1).$$

This lognormal model forms the starting point of the most common option pricing models, especially the famous Black–Scholes model. The dynamic version of this model leads to the well-known stochastic process known as the geometric Brownian motion. In this dynamic model, the random evolution of stock price S_t in time t is described by a *stochastic differential equation*

$$S_t = S_0 + \int_0^t \mu S_s ds + \int_0^t \sigma S_s dW(s), \tag{73}$$

where $W(s)$ is a Brownian motion, $\mu \in \mathbb{R}$, $\sigma \in (0, \infty)$ and the last term on the right side is a *stochastic integral*, that is, integration is performed with respect to a stochastic process. However, it should be noted that lognormal distributions do not fit well the empirical properties of most financial data and time series: much more realistic models can be obtained by using exponential Lévy processes and corresponding more heavy-tailed and possibly skewed distributions such as Student's t, skewed Student's t, variance-gamma or other generalized hyperbolic distributions, or even (what B. Mandelbrot has argued since 1960s) extremely heavy-tailed α-stable distributions.

Shares can be bought either from organized stock exchanges (*listed stocks*[33]) or from so-called OTC (Over-The-Counter) markets (*non-listed stocks*[34]).

[33] listed stocks = noterade aktier = pörssinoteeratut osakkeet

[34] non-listed stocks = onoterade aktier = noteeraamattomat osakkeet

Exchange-traded stocks are usually seen as less risky than stocks trading on the OTC market. This is because both trading in exchanges and corporations with exchange-traded stock are heavily regulated by authorities and subject to stringent requirements concerning timely disclosure of relevant information. These regulations and requirements aim to increase transparency and liquidity, so that the value of the corporation could be assessed reliably and a market price of shares reflecting all relevant information would be available at all times. In contrast, non-listed corporations do not have corresponding incentives and means to timely information disclosure, and hence their evaluation is a more challenging problem. Consequently, a fair price may not be available at all times (or at all). This leads to illiquidity, as the lack of a fair market price reflecting relevant information requires the potential buyer of shares to engage in evaluation procedures, which are costly and time-consuming, and hence increase transaction costs and required processing time before the transaction can be completed. At worst, this may dissuade the buyer from the transaction completely, as the deal may not be considered worth the trouble, or the resources of the buyer are simply not sufficient to carry out such an evaluation procedure.

Naturally, the additional risk of non-listed stocks is usually paired with a higher expected return in comparison with listed stocks.

1.10.5. *Real Estate.* In addition to stocks and bonds, investors can invest in *real estate*[35]. Direct real estate investments are made by purchasing real estate (e.g. mall buildings, apartment buildings). Return on investment is generated by rental income from renting out the property and appreciation of the property value. Rental income is typically a quite stable source of returns, as rents tend to be set contractually for relatively long periods. Rental income can also to some extent act as a hedge against inflation, since rents are often linked to some price index. Appreciation of property value, in contrast, is rather a volatile source of returns.

In terms of riskiness, direct real estate investments are less liquid than stocks or bonds, which typically can be liquidated in a few days at most. In contrast, it may take months or even years to find a buyer for an apartment building or industrial property. So as an investment alternative real estate suffers from illiquidity. Its advantages are usually thought to be lower volatility compared to stocks and higher expected returns in comparison with bonds. Additionally, direct real estate investments are considered to have relatively low correlation with bond and stock market movements – this way, investment in real estate offers diversification benefits as the portfolio is less exposed to adverse market movements on stock and bond markets. Observe that most of these arguments for real estate investments are in the final analysis empirical issues, and there is by no means an unanimous agreement on their validity.

It is possible to invest indirectly on real estate markets by buying shares in a traded real estate fund. In this case, however, some diversification and volatility benefits are likely lost as such fund shares are more volatile and more exposed to stock market movements.

Real estate market is divided into *residential* and *commercial* submarkets, which have slightly different risk and return characteristics. Residential real estate tends to have lower

[35]real estate, property = fastigheter = kiinteistö(sijoitukset)

expected returns but is less exposed to business cycles (people have to live somewhere even during recessions), while commercial real estate, especially industrial and retail real estate, can be very sensitive to economic conditions, which is reflected in higher expected returns.

1.10.6. *Alternative Investments.* Traditional investments are most often considered to consist of bonds and (listed) stocks, with stocks representing the more aggressive (riskier but offering better long-term returns) option. In this presentation, we also consider real estate to be a traditional investment opportunity.

In the past portfolios of institutional investors tended to consist solely of these three types of investment opportunities. In recent years, however, institutional investors such as pension funds have increasingly began to explore other investment opportunities to increase the expected return of their portfolios. Such alternative investments include commodities, hedge funds and private equity.

Commodities

Assets can be divided into *investment assets* and *consumption assets*. Investment assets are held by a significant number of investors solely for investment purposes, while consumption assets are held primarily for consumption. Observe that investment assets may also be held for consumption by some parties, and some parties may hold consumption assets for investment purposes. Stocks and bonds are examples of pure investment assets with no consumption value, but also precious metals (gold and silver) are investment assets, even though they also have a number of industrial uses. In contrast, most other commodities such as copper, tin, oil or pork bellies are consumption assets, even though some investors use them for investment purposes.

As many commodity prices are influenced by factors that are relatively independent of financial markets, commodity prices are often argued to offer diversification benefits in the sense that they do not move in step with financial markets. Of course, while the demand and supply of wheat may not have much to do with business cycles, the demand for industrial metals clearly depends on how the industry using these metals fares. Commodities are also sometimes considered to offer a kind of inflation hedge, as raising commodity prices are one potential source of inflation, in fact sometimes inflation is measured by the price of a basket of commodities.

Direct investment in commodities is cumbersome, as there may be significant storage costs and liquidity is also a problem (while industrial metals might be in high demand during an economic expansion, the demand may drop considerably in a recession); in addition, many commodities are perishable. However, in recent years it has become easier for investors to get commodity exposure via funds specializing in commodity investments or tracking a commodity price index.

Hedge funds

Hedge funds are funds which aim to generate positive returns in all market circumstances, usually by employing many different investment strategies and investing in several different asset classes. This is in contrast to traditional stock or bond funds, which invest mainly only in a single asset class or at most to two or three asset classes and mostly can only take long positions (i.e. no short selling and usually limited opportunities to use leverage). Hedge funds have been able to do short selling and use leverage quite freely, as they have been for the most part relatively unregulated entities. They are somewhat controversial investment opportunities in that while some funds have been able to generate fairly consistently high returns with apparently lower risk (volatility...) than riskiest assets such as stocks, some funds in turn have made spectacular failures, wiping out their investors' value completely and occasionally even causing worldwide disturbances in financial markets in the process (a notorious example is Long Term Capital Management hedge fund in 1998, which had Nobel prize winning financial economists R. Merton and M. Scholes on its payroll).

Hedge fund investments are often illiquid in the sense that the investors' right to redeem their shares is limited, for example by a requirement of advance notification, say, 6 months prior to redemption. They are fairly unregulated entities and often lack transparency: meaning, in essence, that the investor has no way of knowing what the fund is actually doing with the investor's money (cf. the Bernard Madoff case). In addition to increased risk of outright fraud, lack of transparency makes it difficult to assess the actual benefits of a hedge fund investment: two important questions are if the manager has been skillful or just lucky and if the returns are due to strategies uncorrelated with the market or just to exposure to certain markets. Most hedge funds charge high fees, which may not be justified if the returns are mostly based on market exposures which the investor could take directly.

There is some research indicating that the market-independent part of hedge fund returns, so-called alpha, has diminished during the last decade: while hedge funds ten years ago were perhaps able to generate substantial market-independent returns, nowadays much of hedge fund returns consists of market-dependent returns, so-called beta. This could be explained by the increased number of hedge fund managers competing for same scarce alpha opportunities, or weaker average performance due to entrance of less skillful managers into the market.

In any case, the choice of manager plays a fundamental role in hedge fund investing. Both sufficient transparency and liquidity, and furthermore capability to generate consistently high returns by skillful investing should be required from a manager; otherwise the investor likely ends up paying high fees for mediocre investment performance (or even as the victim of a fraudster). It would appear to be the case that only the best funds generate added value sufficient to justify the high management fees.

Private equity

Private equity investments are capital investments in non-listed companies. They are usually made either by entering into a private equity partnership or by buying a share in a fund investing in private equity. In a partnership the investors are usually so-called limited partners who agree to supply capital up to a specified maximum amount to the general partner during a specified period.

Private equity investments are typically highly illiquid: most often the limited partner is obligated to provide the general partner with funds up to a pre-specified maximum amount during the first years of investment period. Possible positive returns are generated usually after the initial period (after 5-8 years) and the limited partner is then entitled to a specified part of these profits. Very often profits are generated by an IPO (Initial Public Offering) of the private equity firm where the firm's stock is listed for public trading at an exchange; limited partners can then either sell their shares at this point or wait for a couple of years more for the share price to appreciate further. However, during the initial period investment cannot usually be realized even if the need should arise.

These concerns may be somewhat alleviated by investing in a private equity fund diversified over different vintages (this refers to the initiation year of the project; returns from private equity investments can be very sensitive to this aspect as for example project begun at the end of a euphoric bull market may be expensive to get into while things still look good in the economy, but may run into serious trouble if things turn around and the economy slides into recession). Naturally, diversification will also likely smooth out the extremely high returns that the most successful private equity ventures can generate.

As with hedge funds, it can be argued that the choice of manager (or in case of a direct private equity investment, the private equity venture) plays a central role. Some research indicates that while the best 20 % of private equity investments can generate abnormally high returns, on average the returns are similar or even lower than those obtained from traded stocks.

1.11. **Financial Derivatives.** *Financial derivatives*[36] are financial instruments whose value is a function of the value of another financial instrument, so-called *underlying*[37]. The underlying instrument can be a stock price, a commodity, an interest rate, an exchange rate, a stock index – or pretty much anything. Derivatives can be used in hedging to reduce risk (and this is, more or less, their original purpose, and is still the most common use of derivative instruments), but they can also be used to make risky bets.

1.11.1. *Forwards and Futures.* A *forward contract*[38] is an agreement where one party agrees to sell (assumes a *short position* in) and the other party agrees to buy (assumes a *long position* in) the underlying instrument (say S) at a specified price (delivery price K) on a specified future moment of time. At the agreement date there are no cash flows between parties (a security may be required). Contract is closed at maturity either by a physical delivery of the underlying or by a cash settlement equal to the difference $S_T - K$ of the spot price at maturity and the delivery price. Forward contracts are traded in over-the-counter markets and are non-standardized, so their details can be negotiated freely by the parties, and there is usually no exchange of payments before maturity. Usually these parties are either two financial institutions or a financial institution and its client.

Forward contracts on foreign exchange are popular tools for hedging currency risk. Suppose that a U. S. corporation knows that it will need to pay 1 million EUR in six months (on July 15, 2010). This payment's value in USD will fluctuate due to EUR/USD exchange rate movements, and may go up or down. If the corporation in question is not willing to

[36](financial) derivative = finansiellt derivat = johdannainen
[37]underlying = underliggande tillgång = kohde-etuus
[38]forward contract = terminavtal = termiinisopimus

bear this exchange rate risk, it may enter into a forward agreement to buy 1 million EUR six months from now at the six month forward exchange rate of (say)1.5021. Then the corporation has assumed a long position and agreed to buy 1 million EUR from the bank for 1.502100 USD on July 16. The bank in turn has assumed a short position and agreed to sell 1 million EUR for 1.5021 million USD. Observe that a forward contract is a binding agreement with no optionality features. The corporation can realize either a loss or a gain from this agreement in relation to the spot exchange rate six months from now – if the rate falls, the corporation realizes a loss in the sense that the 1 million EUR could have been bought from the spot market with lesser amount of USD; if the rate rises, the corporation realizes a gain as it gets 1 million EUR with lesser amount of USD than it would have had to pay if buying EUR from the spot market.

The payoff from a long position in a forward contract on one unit of an asset is $S_T - K$, where S_T is the asset's spot price at delivery date T and K is the agreed-upon delivery price. Correspondingly, payoff from a short position in a forward contract is $K - S_T$. These payoffs can be positive or negative. Entering into a forward contract costs nothing, so the payoff is also the total gain or loss. In pricing forward (and futures) contracts, we make the following assumptions:

- **i:** There are no transaction costs when trading;
- **ii:** All trading profits have the same tax rate;
- **iii:** Money can be borrowed and lent at the same risk-free rate of interest;
- **iv:** Arbitrage opportunities are taken advantage of as they occur.

Forward price is the market price that would be agreed to today for delivery of the asset at a specified future point in time and at a specified price. It is usually different from the spot price and varies with the maturity date. We can derive forward prices for investment assets from arbitrage arguments (for consumption assets, arbitrage arguments yield only lower and upper bounds for the forward price).

We consider first the case where the underlying is a non-dividend paying stock. As it costs nothing to enter into a forward contract, the fair value of the contract at $t = 0$ must be zero. This implies that the forward price $F(t, T)$ must satisfy $F(0, T) = S_0 e^{r_f T}$, where S_0 is the market price of the underlying at time 0, r_f is the risk-free rate of interest and T is the maturity of the forward contract. To why this is so, consider first what would happen if $F(0, T) < K$. Then an arbitrageur could borrow sum S_0 at risk-free rate, buy the stock and take a short position in the forward contract with delivery price K, obtaining a risk-free profit of $K - S_0 e^{r_f T}$. If, on the other hand, $F(0, T) > K$, an arbitrageur could short sell the stock, invest the proceedings of the sale to yield the risk-free rate and take a long position in the forward contract with delivery price K, obtaining a risk-free profit $S_0 e^{r_f T} - K$. Hence, to avoid arbitrage we must have $K = F(0, T)$, i.e. forward price must equal delivery price. Market value of a long position in a forward contract at time t is

$$V_F(t, T) = e^{-r_f (T-t)}(F(t, T) - K),$$

where $F(t, T) = S_t e^{r_f (T-t)}$ is the current forward price at time t (the previous arbitrage argument generalizes trivially to any t). For a short position in a forward contract, we have similarly

$$V_F(t,T) = e^{-r_f(T-t)}(K - F(t,T)).$$

The difference between the forward price and the underlying spot price at time t is called the basis of the forward $basis(t) = S_t - F(t,T)$.

For a dividend paying stock, the price of a forward contract is $F(0,T) = (S_0 - PV_{DIV})e^{r_f T}$; i.e. the present value of dividends to be paid before maturity PV_{DIV} needs to be deducted from the spot price. In case dividends are expressed as a percentage rate q with continuous compounding, the forward rate may be written as $F(0,T) = S_0 e^{(r_f - q)T}$.

For a bond, the price of a forward contract is $F(0,T) = [P_{bond}(0,T) - PV_{CPN}(0,T)]e^{r_f T}$, where $PV_{CPN}(0,T)$ is the present value of coupon payments made before maturity.

Ownership of a physical commodity may provide additional income, which is called convenience yield and denoted by y. If storage costs per unit are also expressed as a percentage rate u with continuous compounding, then the forward price is $F(0,T) = S_0 e^{(r_f + u - y)T}$.

A *futures contract*[39] is similar to a forward contract, but unlike forward contracts, futures contracts are traded on an exchange, and are standardized and marked-to-market daily: if the market price of the contract changes significantly, the parties are required to make additional margin deposits. This marking-to-market mechanism provides a guarantee to the parties that the contract will be honored. The specification of a futures contract by an exchange includes at least specifying the underlying asset, the contract size, where the delivery will be made and when will the delivery be made. Some contracts include several alternatives for example with respect to the delivery location and as a rule the party assuming the short position decides which alternative is used. Unlike forward contracts, which have a specified delivery date, futures contracts typically have a longer delivery period (often a month) during which the party assuming the short position can make delivery.

Futures contracts on a wide variety of financial assets (stock indices, currencies, bonds...) and commodities (pork bellies, live cattle, sugar, wool, lumber, copper, aluminium, gold, tin...)are traded on exchanges such as the Chicago Board of Trade (CBOT), Chicago Mercantile Exchange (CME), London International Financial Futures and Options Exchange and Eurex.

The majority of futures contracts do not lead to delivery, as most traders close out their positions prior to the delivery period by entering into the opposite type of trade from the original one. Then the trader's gain or loss is determined by the change of the futures price between the time of buying the futures contract and the time of closing out the position.

1.11.2. *Swaps.* A swap is an agreement between two parties to exchange streams of cash flows over a specified period in the future. Swaps are usually non-standardized contracts and most often the two parties do not get directly in touch to arrange a swap but use the help of a financial intermediary (such as a bank) to get in contact with each other and complete the swap. Swaps are generally entered into with minimal initial payments.

[39] future = futur = futuuri

The intermediary guarantees the payments to both swap parties and thus takes on the credit risk of the swap agreement. Hence the swap dealer prices the swap transaction in such a way as to provide a return both for its service and for bearing the default risk of the contract. Price may be in the form of an up-front cash payment or of an adjustment to the rate applied to swap payments.

The most commonly encountered swap agreements are interest rate swaps and currency swaps.

Interest Rate Swaps

In a "plain vanilla"interest rate swap, one party agrees to the other party pay cash flows equal to interest at a predetermined fixed rate on a principal for a specified number of years. The other party, in turn, agrees to pay the first party cash flows equal to interest at a specified floating rate on the same principal for the same time period. Thus, in a plain vanilla interest rate swap a fixed rate is exchanged for a floating rate and vice versa.

A plain vanilla swap with maturity T can be characterized as the difference of a fixed rate coupon bond and a floating rate coupon bond, and its value for an intermediary paying fixed and receiving floating at time t is hence given by

$$V_{swap}(t,T) = P_{float}(t,T) - P_{fixed}(t,T), \tag{74}$$

where

$$P_{fixed}(0,T) = \sum_{t=0}^{n-1} \frac{c_{\tau+t/m}}{(1+r_{\tau+t/m})^{\tau+t/m}} + \frac{N}{(1+r_{\tau+(n-1)/m})^{\tau+(n-1)/m}}$$

is the price of a bond paying fixed rate coupons c_t with n outstanding periodic payments m times a year (here r_s is the annual spot rate to maturity s, N is the principal and τ the actual time to the next coupon payment date) and

$$P_{float}(0,T) = \sum_{t=0}^{n-1} \frac{N \cdot \frac{r^{float}_{\tau+t/m}}{m}}{(1+r_{\tau+t/m})^{\tau+t/m}} + \frac{N}{(1+r_{\tau+(n-1)/m})^{\tau+(n-1)/m}}$$

is the price of a bond paying floating rate coupons. The floating rate is always adjusted so that immediately after coupon payment date the price of the floating rate bond equals its principal value. Since the value of the swap is zero when it is first negotiated, the required floating interest rate can be calculated from equation (74). The value depends only on the term structure of interest rates.

Currency Swaps

In a currency swap a loan in a foreign currency is transformed into a loan in the domestic currency. The principal amount must be specified in both currencies and the principal amounts are usually exchanged at the beginning and at the end of the lifetime of the currency swap contract. Principal amounts are chosen to be approximately equal at the exchange rate at the beginning of the contract. If there is no default risk, the value of a currency swap is the difference of two bonds, one denominated in domestic currency and another denominated in foreign currency:

$$V_{swap}(t,T) = S(t) \cdot P_{foreign}(t,T) - P_{domestic}(t,T), \tag{75}$$

where $S(t)$ is the spot exchange rate at time t, expressed as the number of units of domestic currency per unit of foreign currency. Since the value of the swap is zero when it is first negotiated, the required interest rate for the domestic bond can be calculated from the above equation. The value of the swap during the lifetime of the contract depends on the term structure of interest rates in domestic and foreign markets, and on the spot exchange rate.

Credit Default Swaps

Credit derivatives are contracts whose payoff depends on the creditworthiness of one or more commercial or sovereign entities. Most popular of these are *credit default swaps* (CDS). A CDS provides credit insurance against the risk of a default by a particular company, so-called *reference entity.* Default of the company is called a *credit event.* In exchange for periodic payments to the seller during the lifetime of the CDS, the buyer of the CDS obtains the right to sell a particular bond issued by the reference entity (*reference obligation*) to the seller for its par value (the *notional principal* of the swap) if a credit event occurs.

A CDS is settled either by physical delivery or in cash. In physical delivery, the buyer delivers the bonds to the seller in exchange for their par value; in cash settlement, a mid-market price Z of the reference obligation after the credit event is polled from dealers and the buyer receives the difference of the par value and this mid-market price.

To illustrate, suppose that two parties enter into a 5-year CDS with notional principal 100 million euros and the buyer agrees to pay 90 basis points annual payment in exchange for protection against default by reference entity. Now, if the reference entity does not default during the next 5 years, the buyer receives no payoff and pays 900000 euros each year to the protection seller. If there is a credit event, and the contract requires physical delivery, the buyer can sell the bonds to the protection seller for their par value of 100 million euros. If the contract requires cash settlement, and the value of the bonds after the credit event would be 35 % of the original par value, then the buyer would receive 65 million euros in cash settlement. In case the credit event is notified between the periodic payments, the buyer would be required to pay the accrued protection fee for the period between last payment and notification.

A CDS is in some sense a form of credit insurance, however there are some important differences to actual insurance. Firstly, in insurance, the insured asset must be owned by the buyer of protection. In contrast, the buyer of a CDS need not actually own any bonds of the reference entity – a CDS can be used in speculative fashion, to make a bet on a default by the reference entity. If the contract requires cash settlement, bonds need not be bought at all, and in case of a physical delivery, the bonds could in principle be bought from the market after the credit event for immediate delivery. Secondly, insurers are usually required to hold loss reserves, which gives some protection to policyholders.

Issuers of CDSs are not required to hold any special reserves to ensure that the protection can be provided when required.

In pricing of CDSs, important factors are default probabilities, recovery rates and default correlations. Indeed, getting the correlations between defaults wrong can have devastating effects in terms of causing a system-wide undervaluation of default risk.

CDS markets are nowadays large and liquid enough that many market participants infer implied default probabilities from market prices of CDSs.

1.11.3. *Options.* An option is a financial contract which gives its holder (the buyer of the option) the right but not the obligation (i.e. the option) to exercise the option, that is, obtain the payoff specified in the contract, which can be positive or negative, depending on the development of the value of the underlying instrument. If the payoff is negative, the holder will not exercise the option and lets it expire. Due to this (valuable) optionality, unlike with a forward contract, there is a cost to acquiring an option, so-called premium, which the buyer of an option pays to the writer (seller) of the option.

In the case of so-called plain vanilla options the option holder's payoff arises from buying (resp. selling) the underlying instrument from (resp. to) the issuer of the option at a predetermined price (the *strike price*) during some pre-specified exercise period. An option to buy is called a *call option*[40] and an option to sell a *put option*[41]. Observe the asymmetric nature of an option contract's payoffs: for the writer, the maximum profit possible is the received premium (if the option expires unexercised) but the potential loss can be unlimited (this is the case in a call option); for the buyer, the maximum loss possible is the paid premium, but the potential profit can be unlimited.

Options are divided into *styles* according to the degree to which exercise rights of the holder are restricted. A *European option* can be exercised only at maturity of the contract; an *American option* can be exercised at any time before the maturity. These represent the extremities with respect to flexibility with regard to exercise time and are by far the most usual option styles in the markets. Options with styles other than these are often called *exotic options*.

Options are referred to as being *in-the-money*, *at-the-money* and *out-of-the-money* at time t depending on whether an immediate exercise by the buyer at time t would lead to positive, zero or negative cash flow. The intrinsic value of an option is the maximum of zero and the value of the option if exercised immediately (for a call option $\max(S_t - K, 0)$ and for a put option $\max(K - S_t, 0)$).

We will first look at the Black–Scholes option pricing methodology and begin with the baseline case of the underlying being a non-dividend paying stock.

The basic assumptions under which the Black–Scholes formula for pricing of European call options on a stock was derived are

- **i:** Stock price process is a geometric Brownian motion, i.e. the price of the stock at time t is lognormally distributed with parameters μt and $\sigma^2 t$;

[40] call option = köpoption = osto-optio
[41] put option = säljoption = myyntioptio

ii: There are no transaction costs or taxes;
iii: No dividends are paid on the stock during the lifetime of the option;
iv: There are no arbitrage opportunities;
v: Stock trading is continuous;
vi: Money can be borrowed or lent at the same constant risk-free interest rate.

The *Black–Scholes formula* for the price c of a European call option to buy one share of the stock is then

$$c = S \cdot \Phi(d_1) - Ke^{-r_f T}\Phi(d_2), \tag{76}$$

where Φ is the cumulative distribution function of the standard normal distribution and

- $d_1 = \frac{\ln(S/K)+(r_f+\sigma^2/2)T}{\sigma\sqrt{T}}$,
- $d_2 = \frac{\ln(S/K)+(r_f-\sigma^2/2)T}{\sigma\sqrt{T}}$,
- S is the current stock price,
- K is the strike price,
- T is the time to expiration, and
- r_f is the risk-free rate.

The Black–Scholes formula for the price of a put option on the stock can be derived from the pricing formula for a call option using a relationship known as the *put–call parity.* To derive the put–call parity relationship, consider the following two portfolios A and B:

A: One European call option and an amount of cash equal to $Ke^{-r_f T}$, i.e. $c+Ke^{-r_f T}$;
B: One European put option and one share of stock, i.e. $p + S$.

Both portfolios are worth the same at the expiration of the options: for portfolio A the value at expiration of the call is

$$\max(S-K,0)+K = \begin{cases} K & ,S \le K \\ S & ,S > K \end{cases}$$

and for portfolio B the value at expiration of the put is

$$\max(K-S,0)+S = \begin{cases} S & ,S > K \\ K & ,S \le K. \end{cases}$$

Being European, the options cannot be exercised prior to the expiration date. To avoid arbitrage opportunities, the two portfolios must have identical values today, that is

$$c + Ke^{-r_f T} = p + S, \tag{77}$$

and this relationship is called put–call parity. Solving the price of the put from (77) and substituting from (76) for c, we get the Black–Scholes formula for a European put on a stock:

$$p = Ke^{-r_f T} \cdot \Phi(-d_2) - S \cdot \Phi(-d_1).$$

The volatility of the stock price σ is the only parameter in the Black–Scholes formula which cannot be observed directly. It has to be estimated from past stock prices. The Black–Scholes formula can also be used to calculate implied volatilities by observing option prices on the market and solving then the volatility parameter from the formula.

From a risk management point of view, it is important to measure the sensitivity of the value of an option to changes in variables influencing the value. In the Black–Scholes framework, this comes down to the computing the partial derivatives of the option price with respect to the parameters in the formula. The following sensitivities are commonly used for a European call:

- $\Delta_c = \frac{\partial c}{\partial S} = \Phi(d_1)$ measures the rate of change of the value with respect to the stock price,
- $\Gamma_c = \frac{\partial \Delta_c}{\partial S} = \frac{\phi(d_1)}{S\sigma\sqrt{T}}$ measures the rate of change of the option's delta with respect to the stock price,
- $\Theta_c = \frac{\partial c}{\partial T} = \frac{S\phi(d_1)\sigma}{2\sqrt{T}} - r_f K e^{-r_f T}\Phi(d_2)$ measures the rate of change of the value with respect to the expiration time,
- $Vega_c = \frac{\partial c}{\partial \sigma} = S\phi(d_1)\sqrt{T}$ measures the rate of change of the value with respect to the stock volatility, and
- $Rho_c = \frac{\partial c}{\partial r_f} = KTe^{-r_f T} \cdot \Phi(d_2)$ measures the rate of change of the value with respect to the risk-free rate.

Similar measures can be derived for European put options.

It is clear that a European call becomes more valuable as the stock price increases and less valuable as the strike price increases, while for a European put option the opposite is the case.

Increased volatility increases the probability of large price movements in either direction. However, since the buyer of an option has limited downside risk and unlimited upside risk, for both calls and puts increase in volatility increases the value of the option.

As the risk-free rate increases, the present value of future cash flows decreases. For the holder of a call option, the present value of the possible future payment (strike price) to be made decreases and hence the value of the option increases; for the holder of a put option, the present value of the possible future payment (strike price) received decreases and hence the value of the option decreases.

The time to expiration influences the option value through interest rate and volatility. A European call's value will decline as time passes, since the present value of the potential future payment increases and the volatility decreases. For a European put, the interest rate effect and the volatility effect have opposite impacts and so the impact of the time to expiration to the value of the option is ambiguous.

Observe that sensitivities are calculated assuming that all other variables remain fixed, which is not the case in practice.

Previous discussion concerned only European options which can be exercised only at expiration date. As stated in the beginning of this section, *American options* can be exercised at any time before expiration, and hence they contain more optionality than

European options. Thus the value of an *American call option*, C, must be at least as large as the value of a European option that expires somewhere in between:

$$C \geq \max(c_{t_0}, c_{t_1}, \ldots, c_{t_n=T}) \text{ and } P \geq \max(p_{t_0}, p_{t_1}, \ldots, p_{t_n=T}).$$

So $C \geq c$. The price of a European call option on a non-dividend paying stock decreases as time to expiration decreases, which would imply that $C \geq c_T = c$. Early exercise of an American call option on non-dividend paying stock kills the time value of the option. From put–call parity we have that

$$c = p + S - Ke^{-r_f T} \geq \max(S - Ke^{-r_f T}, 0) \geq \max(S - K, 0),$$

where $\max(S-K, 0)$ is the payoff of an American call option if exercised. This shows that $c \geq C$ and it is never optimal to exercise an American call prior to maturity.

For *American put options*, to exercise prior to maturity may be optimal if the option is deep in-the-money as then the gain from losing less interest income can outweigh the cost of giving up the upside potential of the volatility.

In general, pricing of American options requires solving an *optimal stopping problem*

$$\max_{\tau \in \mathcal{T}_T} \mathbb{E}\left[e^{-r_f \tau} g(S_\tau)\right],$$

where $\mathcal{T}_T$ is a set of suitable so-called stopping times $\tau \leq T$, T is the expiration time and $g(x)$ is the payoff of the option as a function of the price of the underlying asset. The stopping time $\tau^\star$ maximizing the above expression is the optimal time to exercise and the fair price of the option is then $\mathbb{E}\left[e^{-r_f \tau^\star} g(S_{\tau^\star})\right]$. In this framework the underlying asset price S is assumed to follow some suitable stochastic process, which could be a geometric Brownian motion, a linear diffusion, an exponential Lévy process, a Lévy (jump) diffusion or some more general process.

The Black–Scholes formula can be adjusted to deal with an underlying other than a non-dividend paying stock by adjusting the spot price S suitably. For a *dividend paying stock*, we set

$$S \leftarrow S - PV_{DIV},$$

where PV_{DIV} is the present value of dividends that will be paid until the expiration. Observe that in this case an early exercise of an American call option may be optimal if the value of dividends is sufficiently large, since the value of the stock is expected to decrease after the dividend payment.

For a *stock index paying a continuous dividend yield* q, we set

$$S \leftarrow Se^{-qT}.$$

For a *currency option*, S is the current spot price in domestic currency of one unit of foreign currency. The foreign currency holder receives a benefit equal to risk-free rate on the foreign currency; this is comparable to a continuous dividend yield and so we set

$$S \leftarrow Se^{-r_f^{foreign} T},$$

where $r_f^{foreign}$ is the risk-free interest rate on the foreign currency.

An *option on futures* is the right, but not the obligation, to enter into a futures contract. Hence we set

$$S \leftarrow Fe^{-r_f T},$$

where F is the current price of the futures contract.

The Black–Scholes formula can be applied to European calls and puts and American calls whose price is equal to the price of the European call. However, this approach cannot be applied to the American put option or other more complex options. Another approach to valuing options is the so-called *binomial model* or *Cox–Ross–Rubinstein (CRR) model*, in which it is assumed that the future evolution of the price of the underlying can be described by a binomial tree: at each time step, the value can "go up"with probability ψ or "go down"with probability $1-\psi$.

In a one-step binomial model, the future spot price

$$\tilde{S}_T = \begin{cases} S_0 e^{uT} & \text{, with probability } \psi \\ S_0 e^{dT} & \text{, with probability } 1-\psi, \end{cases}$$

where u, d and ψ are given parameters satisfying $d \leq r_f \leq u$. The option is defined as a function $f(S)$ describing the payoff of the option as a function of the current price S of the underlying. Consider now a portfolio consisting of Δ units of the underlying asset and a short position in one option. The value of this portfolio at maturity T is

$$\Delta S_0 e^{uT} - f(S_0 e^{uT}),$$

if the asset price goes up, and

$$\Delta S_0 e^{dT} - f(S_0 e^{dT}),$$

if the asset price goes down. The portfolio is deterministic (i.e. risk-free) if these values are equal, that is, if

$$\Delta S_0 e^{uT} - f(S_0 e^{uT}) = \Delta S_0 e^{dT} - f(S_0 e^{dT}) \Leftrightarrow \Delta = \frac{f(S_0 e^{uT}) - f(S_0 e^{dT})}{S_0 e^{uT} - S_0 e^{dT}}.$$

So for the portfolio to be risk-free, delta must equal the ratio of the change in the option price to the change in the price of the underlying asset. A risk-free portfolio must earn the risk-free rate of interest, and the cost of setting up this portfolio is $\Delta S_0 - P$, where P is the price of the option. Thus we must have that

$$[\Delta S_0 e^{uT} - f(S_0 e^{uT})] = e^{r_f T}(\Delta S_0 - P).$$

Inserting the expression for delta and simplifying yields that the price of the option is

$$P = e^{-r_f T}[\psi \cdot f(S_0 e^{uT}) + (1-\psi) \cdot f(S_0 e^{dT})],$$

where

$$\psi = \frac{e^{r_f T} - e^{dT}}{e^{uT} - e^{dT}}.$$

That is, the price of the option is the expected present value of the payoff under a *risk-neutral probability measure* $\mathbb{Q}$ such that

$$\mathbb{Q}(\{S_T = S_0 e^{dT}\} = 1 - \psi \text{ and } \mathbb{Q}(\{S_T = S_0 e^{uT}\} = \psi.$$

The expected stock price at time T under the risk-neutral probability measure $\mathbb{Q}$ is

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}[\tilde{S}_T] &= \psi S_0 e^{uT} + (1-\psi) S_0 e^{dT} \\ &= \psi S_0 (e^{uT} - e^{dT}) + S_0 e^{dT} \\ &= S_0 (e^{r_f T} - e^{dT}) + S_0 e^{dT} \\ &= S_0 e^{r_f T},\end{aligned}$$

that is, under the risk-neutral probability measure, the stock price is expected to increase at the risk-free rate of interest.

A two-step binomial model can be constructed similarly, leading to a tree structure

$$\begin{array}{ccccc} & & & \to & S_0 e^{(u+u)h} \\ & \to & S_0 e^{uh} & & \\ S_0 & & & \to & S_0 e^{(u+d)h} \\ & \to & S_0 e^{dh} & & \\ & & & \to & S_0 e^{(d+d)h} \end{array}$$

where the time step $h = T/2$ and $d \leq r_f \leq u$. At time T

$$\tilde{S}_T = \begin{cases} S_0 e^{2uh} & \text{with probability } \psi^2 \\ S_0 e^{(u+d)h} & \text{with probability } 2\psi(1-\psi) \\ S_0 e^{2dh} & \text{with probability } (1-\psi)^2, \end{cases}$$

where ψ is the probability of an increase in the price of the underlying asset. In similar fashion, we can construct n-step binomial models, where the pricing formula is

$$P = e^{-r_f T} \sum_{k=0}^{n} \binom{n}{k} \psi^k (1-\psi)^{n-k} \cdot f(S_0 e^{[ku+(n-k)d]\,h}).$$

Binomial models can be used to value American options and more complicated options, however in the case of American options the value function f must be adjusted at each time step to take into account the possibility of an early exercise.

Example: (*Two-step binomial model for a put option*) Let $T = 1$, $n = 2$, $h = 1/2$, $S_0 = 100$, $K = 95$, $r_f = 10$ %, $e^{uh} = 1.1$ and $e^{dh} = 0.9$. Then the binomial tree is

$$\begin{array}{ccccc} & & & \to & S_0 e^{(u+u)h} = 121 \\ & \to & S_0 e^{uh} = 110 & & \\ S_0 = 100 & & & \to & S_0 e^{(u+d)h} = 99 \\ & \to & S_0 e^{dh} = 90 & & \\ & & & \to & S_0 e^{(d+d)h} = 81 \end{array}$$

and the one step up probability is

$$\psi = \frac{e^{r_f h} - e^{dh}}{e^{uh} - e^{dh}} = 0.7564.$$

The payout at maturity is

$$f(S_T) = \max(K - S_T, 0) = \begin{cases} 0 & , S_T = 121 \\ 0 & , S_T = 99 \\ 14 & , S_T = 81. \end{cases}$$

If the option is European, then the expected payoffs at first time step are

$$f(S_0 e^{uh}) = e^{-r_f h} \left[\psi f(S_0 e^{2uh}) + (1-\psi) f(S_0 e^{(u+d)h})\right] = 0$$

$$f(S_0 e^{dh}) = e^{-r_f h} \left[\psi f(S_0 e^{(u+d)h}) + (1-\psi) f(S_0 e^{2dh})\right] = 3.2467$$

and the price of the European put is

$$p = e^{-r_f h} \left[\psi f(S_0 e^{uh}) + (1-\psi) f(S_0 e^{dh})\right] = 0.7520.$$

If the option is American, we must take into account the value of early exercise by making the transformation

$$f(S_0 e^{uh}) \leftarrow \max\left(f(S_0 e^{uh}), K - S_0 e^{uh}\right) = \max(0, -15) = 0$$

$$f(S_0 e^{dh}) \leftarrow \max\left(f(S_0 e^{dh}), K - S_0 e^{dh}\right) = \max(3.2467, 5) = 5.$$

The price of the American put is

$$P = e^{-r_f h} \left[\psi f(S_0 e^{uh}) + (1-\psi) f(S_0 e^{dh})\right] = 1.1589.$$

It is possible to show that for a suitable choice of parameters, the n-step binomial model can be considered as a discrete time version of the Black–Scholes model: as the number of time steps n goes to infinity, the binomial model approaches the Black–Scholes model (basically this is due to the fact that the binomial distribution converges to a normal distribution as $n \to \infty$ which can be proved using a version of the central limit theorem). So in this sense the two pricing methodologies are consistent with each other.

Previously we mentioned that some bonds have embedded options. If these options can be priced, then the price of such a bond is the sum of the prices of the corresponding option-free bond and the price of the option. For example, in a callable bond, the issuer has the right to call the bond. This means effectively that the buyer of the callable bond has bought a noncallable (i.e. usual option-free) bond and sold to the issuer a call option on this bond. The fair price of the callable bond is then

$$P_{CB} = P_{NCB} - P_c,$$

where P_{NCB} is the price of the option-free bond and P_c is the price of a call option on the bond.

1.12. **Some Basics of Investment Portfolio Analysis.**

1.12.1. *Utility and Risk Aversion.* In utility theory it is assumed that the investor's preferences towards risk can be described using a utility function. One way of modelling how investors make decisions under uncertainty is to assume that they act to maximize the expected value of their utility function (expected utility approach).

It is assumed that investors (i) prefer more wealth to less wealth and that (ii) they are risk averse, that is, they do not like risk. To formalize this intuition, we will assume that the investor's utility function u satisfies the following two properties:

i: u is strictly increasing, and
ii: $u(x+t) - u(x)$ is strictly decreasing in x for all $t > 0$.

It can be shown that then u is continuous and strictly concave. Curvature of the utility function reflects the investor's risk preferences: concavity corresponds to risk aversion, linearity to risk neutrality and convexity to risk seeking. For a concave u

$$u\left(\alpha x + (1-\alpha)y\right) \geq \alpha u(x) + (1-\alpha)u(y)$$

for any $\alpha \in [0,1]$ and any x and y in the domain of u. In particular, this implies that

$$u\left(\mathbb{E}[\tilde{X}]\right) \geq \mathbb{E}\left[u(\tilde{X})\right],$$

i.e. in expected utility sense, the certain (risk-free) wealth $\mathbb{E}[\tilde{X}]$ is preferred to uncertain (risky) wealth $\tilde{X}$ which has the same expected value $\mathbb{E}[\tilde{X}]$. A certain wealth W satisfying

$$u(W) = \mathbb{E}[u(\tilde{X})]$$

is called the *certainty equivalent* of the random wealth $\tilde{X}$.

Observe that expected utility functions are used to *rank* alternative random wealth levels – the actual numerical value of utility has no meaning by itself. Utility functions leading to identical rankings are called *equivalent*. In general, given a utility function $u(x)$, any function $v(x)$ of form

$$v(x) = au(x) + b$$

with $a > 0$ is a utility function equivalent to $u(x)$. The certainty equivalent of a random variable is the same for all equivalent utility functions.

Investor's risk aversion, as described by the curvature of his/her utility function, can be measured by the *Arrow–Pratt coefficient of absolute risk aversion*

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

or by the *Arrow–Pratt coefficient of relative risk aversion*

$$R_R(x) := -\frac{xu''(x)}{u'(x)}.$$

Coefficients reflect the strength of the curvature and are normalized by dividing with $u'(x)$ so that they have the same value for all equivalent utility functions.

Some commonly used utility functions are

i: *Exponential* $u(x) = -e^{-ax}$ with $a > 0$;

ii: *Logarithmic* $u(x) = \ln(x)$, which is defined only for $x > 0$ and is extremely risk averse for low levels of wealth (and if there is a positive probability for wealth 0, expected utility will be $-\infty$);

iii: *Power* $u(x) = bx^b$ for $b < 1$, $b \neq 0$. If $b = 1$, we obtain the risk-neutral linear utility function;

iv: *Quadratic* $u(x) = x - bx^2$ with $b > 0$, which is increasing only for $x < 1/(2b)$.

1.12.2. *Portfolio Theory.* Suppose that an investor with preferences described by a non-decreasing, strictly concave and continuously differentiable utility function u has initial wealth v_0 and she can invest in n risky assets whose stochastic returns are given by the random vector $\tilde{r} = (\tilde{r}_1, \ldots, \tilde{r}_n)$. Investor's weights in different assets are

$$w = (w_1, \ldots, w_n), \ \sum_{i=1}^{n} w_i = 1,$$

where w_i is the proportion of wealth invested in asset i. We call w the investor's *portfolio*, and the stochastic return of the portfolio is then $\tilde{r}_p := w^T \tilde{r} = \sum_{i=1}^{n} w_i \tilde{r}_i$ and the stochastic final wealth of the investor is $\tilde{v}_1 = v_0(1 + \tilde{r}_p)$. The investor's optimal portfolio choice $w^\star$ is the solution of the expected utility maximization problem

$$\max_{w \in \mathbb{R}^n} \left\{ \mathbb{E} \left\{ u \left(v_0 (1 + \tilde{r}_p) \right) \right\} \;\middle|\; \sum_{i=1}^{n} w_i = 1 \right\}.$$

Investor's portfolio w has expected return

$$\mathbb{E}[\tilde{r}_p] = \sum_{i=1}^{n} w_i \mathbb{E}[\tilde{r}_i] = w^T \bar{r},$$

where $\bar{r} := \mathbb{E}[\tilde{r}]$, and variance

$$\sigma_p^2 := \sum_{i,j}^{n} w_i w_j \sigma_{ij} = w^T \Sigma w,$$

where $\Sigma = cov(\tilde{r})$, the covariance matrix of random vector $\tilde{r}$.

A portfolio x is *mean-variance efficient*, if there exists no other portfolio y with a higher or equal expected return than x and a lower variance than x. In general, the optimal solution to the investor's utility maximization problem is mean-variance efficient only in two cases:

i: if utility function is quadratic; or

ii: if asset returns have multivariate normal distribution.

It can be shown that the variance of the return of a portfolio can be reduced by including additional assets in the portfolio, a process referred to as *diversification*. How large a reduction can be achieved depends on the interdependencies between returns (we will look at this a bit more closely in the exercises). However, diversification may reduce the overall expected return and hence may not be desirable if much expected return is sacrificed for a small decrease in variability. This is the motivation behind the mean–variance approach developed by H. Markowitz: to make the trade-off between mean and variance explicit.

1.12.3. *Markowitz Model and Capital Asset Pricing Model.* In the setup of the previous subsection, the Markowitz problem is

$$\begin{aligned} &\min \quad \tfrac{1}{2}\textstyle\sum_{i,j=1}^{n} w_i w_j \sigma_{ij} = \tfrac{1}{2} w^T \Sigma w \\ &s.t. \quad \textstyle\sum_{i=1}^{n} w_i \bar{r}_i = w^T \bar{r} = \alpha \\ &\qquad\;\; \textstyle\sum_{i=1}^{n} w_i = w^T \mathbf{1} = 1 \end{aligned} \tag{78}$$

where α is a specified target return. Hence the solution of the problem determines a minimum variance portfolio for a given rate of return. Observe that all assets are here assumed to be risky (there is no risk-free alternative available) and negative weights w_i are allowed – i.e. short selling of assets is permitted. This optimization problem can be solved using the Lagrange method (we omit the details as that is just standard equality constrained optimization), yielding that the n portfolio weights w_i and the Lagrange multipliers λ and μ corresponding to the efficient portfolio having mean rate of return α satisfy the linear equations

$$\begin{aligned} \textstyle\sum_{j=1}^{n} \sigma_{ij} w_j - \lambda \bar{r}_i - \mu &= 0, \text{ for } i = 1, \ldots, n \\ \textstyle\sum_{i=1}^{n} w_i \bar{r}_i &= \alpha \\ \textstyle\sum_{i=1}^{n} w_i &= 1. \end{aligned}$$

These are $n+2$ linear equations in $n+2$ (weights and the two Lagrange multipliers) unknowns, and can be solved with linear algebra methods.

In the previous formulation it was assumed that short selling is allowed. This is not always the case in practice, but instead often weights are required to be non-negative. This adds n non-negativity constraints $w_i \geq 0$ to the Markowitz problem, which then becomes a *quadratic programming problem* with both equality and *inequality constraints*. Such problems cannot in general be reduced to linear equations, but are usually solved with spreadsheet programs or special purpose computer programs. Important difference between the two formulations is that usually when short selling is allowed, most weights have nonzero values. With non-negativity constraints, typically many weights are equal to zero.

Markowitz approach offers the investor tools to either maximize return for a given risk (variance) tolerance or minimize risk for a given desired expected return. But while the idea of characterizing investment opportunities via risk and return is intuitive, it should be observed that using standard deviation as a measure of risk, as Markowitzian theory does, is questionable when underlying statistical distributions are non-normal. While normal distribution is symmetric and characterized completely by its first two moments, i.e. mean and variance, distributions more consistent with empirical observations of financial variables do not have these properties. Standard deviation and variance as measures of riskiness do not take into account possible skewness or heavy tails of the underlying distributions, and, as stated previously, financial variables display very often these features. Moreover, in practice Markowitz approach often suffers from extreme sensitivity to estimated parameters: a small change in parameters may cause huge changes in the optimal portfolio allocation prescribed by mean-variance optimization. As in actual reality parameters used in optimization are some statistical estimates of the unknown true parameters and contain estimation error, this is a big problem. A related problem from practitioner's point of view is the tendency of mean-variance optimization to generate corner solutions, in which only few of the available investment opportunities are used. This is in contradiction to intuition about benefits of diversification.

Because the equations determining the solution of the Markowitz problem are linear, it can be shown that if w^1 and w^2 are solutions of the Markowitz problem with mean returns $\bar{r}^1$ and $\bar{r}^2$, then for any $a \in \mathbb{R}$ the linear combination $aw^1 + (1-a)w^2$ is also a solution of the Markowitz problem with mean return $a\bar{r}^1 + (1-a)\bar{r}^2$. This implies the following *two-fund theorem.*

> **Theorem:** Two efficient portfolios can be established so that any efficient portfolio can be duplicated (in terms of mean and variance) as a combination of these two.

This theorem states that two mutual funds could provide a complete investment service for everyone. This conclusion is, however, based on the premises that everyone cares only about means and variances and has the same assessments of the means and covariance structures, and that a single-period framework is appropriate. All these assumptions are tenuous at best.

Computationally, the theorem implies that solving the Markowitz problem for all possible target returns can be done by finding two solutions and forming combinations of these two. A simple way to get two solutions is to specify the Lagrange multipliers: convenient choices are (1) $\lambda = 0$, $\mu = 1$ and (2) $\lambda = 1$, $\mu = 0$. Solutions thus obtained may violate the constraint $w^T\mathbf{1} = 1$ but this can be remedied by scaling all w_i by dividing the value of each individual w_i with the sum of all w_i's.

A risk-free asset has $\sigma = 0$, that is, its return is known with certainty and has to equal the risk-free rate by definition, and it corresponds to lending (positive weight) or borrowing (negative weight) at the risk-free rate. Availability of a risk-free asset has a profound impact on the set of efficient portfolios, summarized by the following *one-fund theorem.*

> **Theorem:** There is a single fund F of risky assets such that any efficient portfolio can be constructed as a combination of F and the risk-free asset.

To get an idea why this is true, consider an investor able to invest in any efficient portfolio of risky assets and in the risk-free asset. If proportion x_0 is invested in the risk-free asset, then the mean and variance of this portfolio are

$$\mu_p = x_0 r_f + (1-x_0)\mu \text{ and } \sigma_p^2 = (1-x_0)^2\sigma^2,$$

where μ and σ are the expected return and standard deviation of the risky portfolio. In the (μ, σ) diagram, this portfolio lies on a straight line connecting the risk-free asset and the risky portfolio:

$$\mu_p = r_f + \left(\frac{\mu - r_f}{\sigma}\right)\sigma_p.$$

Capital Asset Pricing Model (CAPM) deduces the correct price of a risky asset within the framework of the mean–variance setting and follows logically from the Markowitz portfolio theory. To be more precise, assume the following:

i: all investors are mean–variance optimizers;
ii: all investors agree on the probabilistic structure of asset returns (means, variances and covariances);
iii: there is a unique universal risk-free rate for borrowing and lending;
iv: there are no transaction costs.

By the one-fund theorem, we know, firstly, that given these assumptions, every investor will purchase a single mean–variance efficient fund of risky assets and possibly additionally borrow or lend at the risk-free rate. Secondly, since everyone uses the same means and covariances, everyone will purchase the *same* fund. The mix of the risky portfolio and the risk-free asset will vary depending on the risk preferences of each individual investor, but the risky portfolio is the same for all investors. Hence the one fund of the one-fund theorem is actually the only fund that is used by investors. By an equilibrium argument, this fund must be then equal to the *market portfolio*, the summation of all assets (this equilibrium argument has its strengths and weaknesses, but we will not go into details here).

The set of mean–variance efficient portfolios consists of a single straight line in the $(\bar{r}, \sigma)$ diagram emanating from the point $(r_f, 0)$ and passing through the market portfolio point $(\bar{r}_M, \sigma_M)$. This line is called the *capital market line* and it shows the relation between the expected rate of return and the risk of return (as measured by the standard deviation) for mean–variance efficient portfolios:

$$\bar{r} = r_f + \left(\frac{\bar{r}_M - r_f}{\sigma_M}\right)\sigma.$$

The slope of the capital market line $\frac{\bar{r}_M - r_f}{\sigma_M}$ is called the *price of risk* and it tells by how much the expected rate of return must increase if the standard deviation of the rate increases by one unit.

Capital Asset Pricing Model shows how the expected rate of return of an individual asset relates to its individual risk. It states that if the market portfolio is efficient, then the expected return $\bar{r}_i$ of any asset i satisfies

$$\bar{r}_i - r_f = \beta_i(\bar{r}_M - r_f), \tag{79}$$

where $\beta_i = \frac{\sigma_{iM}}{\sigma_M^2}$. Here the expected excess return of asset i over the risk-free rate is a multiple of the excess return of the market portfolio over the risk-free rate, where the multiplier is the asset's β, which is the normalized covariance of the asset with the market portfolio. Hence it is the *covariance with the market* which determines the expected excess return of the asset. While we still measure the overall risk of a portfolio in terms of standard deviation, we do not care about the individual standard deviations of assets but instead consider their β's. To illustrate, if an asset has zero (or negative) β, the expected rate of return for that asset is the risk-free rate (or lower than the risk-free rate). This is because even if the rate of return of a $0-\beta$ asset has a very high variance, as it is uncorrelated with the market the risk can be diversified away completely. An asset with a rate of return negatively correlated with the market rate of return, in turn, can even be used to reduce the overall risk of a portfolio – investors hence can accept a negative risk premium.

CAPM can be represented explicitly as a pricing model as follows. Denote the purchase price of an asset by P and the payoff from selling the asset at some specified later date by $\tilde{Q}$. The (stochastic) rate of return from this investment is then $\tilde{r} = (\tilde{Q} - P)/P$. Replacing $\tilde{Q}$ with its expectation $\bar{Q}$ and plugging the expected rate of return into the CAPM equation we get

$$\frac{\bar{Q} - P}{P} = r_f + \beta(\bar{r}_M - r_f),$$

from which we can solve the price P to get the price of the asset according to CAPM:

$$P = \frac{\bar{Q}}{1 + r_f + \beta(\bar{r}_M - r_f)}. \tag{80}$$

The pricing formula generalizes the familiar discounting principle from deterministic to stochastic situations: if the payoff $\tilde{Q}$ were deterministic, then $\beta = 0$ and the price is just the (certain) payoff discounted with the risk-free rate, but if the payoff $\tilde{Q}$ is genuinely stochastic (random), then a *risk-adjusted* discount rate $r_f + \beta(\bar{r}_M - r_f)$ is used in discounting the expected value of the payoff.

CAPM can be seen as a kind of special case of a *factor model* for asset returns. These are constructed in order to reduce the dimension of the pricing problem (number of covariances to estimate grows rapidly as the number of assets increases, rendering mean–variance optimization computationally unwieldy) . It is assumed that a (small) set of common basic underlying sources of randomness, factors, explains most of the variation in a (large) number of asset returns. Typical factors could be market indices or general economic variables such as industrial production, GDP growth or unemployment rate. An m-factor model for n asset rates of return $\tilde{r}_i$ is of form

$$\tilde{r}_i = a_i + \sum_{j=1}^{m} b_{ij}\tilde{F}_j + \tilde{\epsilon}_i, \; i = 1, \ldots, n,$$

where a_i is a constant, b_{ij} are constants called *factor loadings*, $\tilde{F}_j$ are the factors and $\tilde{\epsilon}_i$ is a stochastic error term uncorrelated with each other and the factors (naturally $m < n$ usually). CAPM can be considered as a one-factor model with the (excess rate of return of) the market portfolio $\bar{r}_M$ as the single factor: taking expectations we get

$$\bar{r}_i - r_f = \alpha_i + \beta_i(\bar{r}_M - r_f),$$

since it can be shown that fitting the single factor model with least squares method yields the factor loadings $b_i = \frac{\sigma_{iM}}{\sigma_M^2} = \beta_i$ (we skip the details here). Difference to the CAPM equation is the constant α_i: CAPM predicts that this should equal zero and from that viewpoint α_i is a measure of the amount that asset i is mispriced. A stock with a positive α_i

is “performing better than it should“– in this context, this quantity is used for performance measurement and called *Jensen's alpha*. Observe, however, that the CAPM model assumes that the market is efficient, while the single-factor model does not make such an assumption – hence these models are not equivalent.

1.12.4. *Portfolio Performance Measurement.* Performance measurement provides an objective quantitative assessment of the change in value of an investment portfolio over a specified time period.

The most common performance measure is the *rate of return (ROR)*, which expresses the gain or loss (earnings) on the portfolio during a period T as a percentage of the amount invested:

$$ROR = \frac{Earnings}{Amount\, invested}$$

or expressed as an annual rate of return r_p

$$r_p = (1 + ROR)^{\frac{1}{T}} - 1.$$

Earnings are valued at market value and include all capital gains (realized and unrealized) and all dividends and interest income (whether reinvested or not). Amount invested is the market value of assets which produced the earnings in the beginning of the period. However, if there is new money added to the portfolio or withdrawals of money from the portfolio during the period, ROR is not a very accurate measure.

Exact rate of return is more accurate measure of portfolio performance. It is calculated by valuing the portfolio each time a cash flow occurs, just before the cash flow is reflected in the portfolio value, calculating the rate of return for each subperiod between individual cash flows and linking the computed rate of return together with all earlier subperiod rates of return; portfolio value after the flow is used as the initial value of the next subperiod. If the time period T is divided into n subperiods by occurrence of cash flows, the exact rate of return is given by

$$EROR_T = (1 + EROR_1)(1 + EROR_2) \cdots (1 + EROR_n) - 1$$

and the annualized return of the portfolio is then

$$r_p = (1 + EROR_T)^{\frac{1}{T}} - 1.$$

Observe that here each period's rate of return is of equal weight, regardless of how much is invested or the length of the subperiod.

In *time-weighted rate of return* the cash flows are weighted based on the actual time they are held in the portfolio. The formula for one subperiod rate of return is

$$TWROR = \frac{MV(end) - MV(beginning) - Net\, cash\, flows}{MV(beginning) + Weighted\, cash\, flows},$$

where $MV(x)$ refers to market value of the portfolio at time x, Net cash flows equals the sum of all positive and negative cash flows during the subperiod, Weighted net cash flows equals the weighted sum of all positive and negative cash flows during the subperiod with each weight equal to the fraction of subperiod remaining after the occurrence of the associated cash flow. The time weighted rate of return for period T divided into n subperiods is then

$$TWROR_T = (1 + TWROR_1)(1 + TWROR_2) \cdots (1 + TWROR_n) - 1$$

and the annualized return of the portfolio is

$$r_p = (1 + TWROR_T)^{\frac{1}{T}} - 1.$$

Return attribution usually refers to explaining the difference between the portfolio return and some benchmark return as being due to some specified factors (which may vary depending on the investor's aims). For example, with stock investments attribution could decompose the return performance into *benchmark performance* and *manager controlled performance*; the latter could further be subdivided into performance contribution due to *security selection*, performance contribution due to *allocation* (deviation from the benchmark allocation) and performance contribution due to *interaction* between selection and allocation. Return attribution is important in evaluating whether an actively managed fund adds value (or destroys value) compared to a passive index tracking investment – an actively managed fund is not a worthwhile investment unless it outperforms its benchmark by more than the cost of management fees. It is also important to understand how different factors contribute to performance – for example, whether outperformance is due to superior security selection or superior allocation decisions of the manager.

1.12.5. *Portfolio Risk Measurement.* In Markowitzian portfolio theory risk is identified with variability of returns, summarized by the standard deviation of observed returns

$$\sigma_p = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(r_i - \bar{r})^2},$$

where n is the number of observations, r_i the i'th observation and $\bar{r}$ the average of observations. This measures the dispersion of returns around their average and is one way of quantifying uncertainty. However, it has several drawbacks: it is a symmetrical measure, while empirical return distributions are asymmetrical (skewed) and investor's attitudes towards large profits or large losses are different; it measures dispersion around the portfolio's own average, making comparisons between portfolios difficult; and historical sample standard deviations may not be a good indication of future values of standard deviations (returns may well be heteroskedastic, i.e. their variance may change with time).

In the context of the Capital Asset Pricing Model, the risk-adjusted comparison of different portfolios is often performed using the *Sharpe ratio*, which is defined as the excess return of the portfolio over the risk-free rate of return, $\bar{r}_p - r_f$, divided by the standard deviation of the portfolio σ_p:

$$S = \frac{\bar{r}_p - r_f}{\sigma_p}.$$

The higher the ratio, the better the risk-adjusted return of the portfolio. Sharpe ratio is used to compare different portfolios to each other or to the market portfolio – in case of using the market portfolio as the benchmark, Sharpe ratio higher than the market portfolio's

Sharpe ratio implies that the portfolio in question offers better risk-adjusted returns than the market. Observe that also Sharpe ratio equates risk with standard deviation and hence disregards the impact of higher moments (skewness, kurtosis) on the risk. In particular, tail risks are not taken into account.

To better accommodate features such as the asymmetry of return distributions, several alternative measures of risk have been developed. One is the so-called *downside risk* $\tilde{\sigma}$, which introduces a target rate of return, so-called *minimum acceptable return* r_{min}, and considers only returns underperforming this rate as contributing to risk:

$$\tilde{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\min(0, r_i - r_{min}))^2}.$$

Observe that while n is the number of all observations, only observations with $r_i < r_{min}$ create nonzero terms in the sum. An analogue of the Sharpe ratio using the downside risk is the *Sortino ratio*

$$R_{Sortino} = \frac{\bar{r}_p - r}{\tilde{\sigma}}. \tag{81}$$

Yet another modification of the Sharpe ratio is *Treynor ratio*

$$R_{Treynor} = \frac{\bar{r}_p - r}{\beta_p},$$

where the risk is (according to the CAPM) identified with the β of the portfolio.

All of the above measures rely essentially on the first and second moments of the distribution of returns and hence do not reflect for example tail risks. Tail risk is often measured with so-called *Value at Risk (VaR)*, which is defined as the α-quantile of the return distribution:

$$VaR_\alpha(\tilde{r}) = \inf\{r \,|\, \mathbb{P}(\tilde{r} \le r) < \alpha\}.$$

So VaR is the maximum loss with probability α: with probability $1-\alpha$ our loss is less than the VaR. Typically α is small so that $1-\alpha$ equals 95 %, 97.5 % or 99.5 % (this is the level used in Solvency II where the solvency capital requirement (SCR) is set equal to one-year VaR at confidence level 99.5 % – in other words, the SCR suffices to cover losses during a year with a 99.5 % probability, or put differently, SCR is expected to be insufficient to cover losses once every 200 years). Problem with VaR (and all measures of tail risk, which relates to rare extreme events) is that we have almost by definition very little data or knowledge about the far tails of the distribution, so estimation is problematic.

VaR has certain drawbacks relevant from both theoretical and practical viewpoints. For non-elliptical distributions, VaR is not a so-called *coherent risk measure*. This relates to the fact that VaR does not take into account the shape of the distribution in the tail beyond the VaR value: VaR does not tell what to expect if a loss larger than VaR is realized: will the expected loss be close to VaR or maybe 20 times VaR? A better measure taking also the tail beyond VaR into account is the *Conditional Value at Risk (CVaR/TailVaR/Expected Shortfall)*

$$CVaR_\alpha(\tilde{r}) = \mathbb{E}\left\{\tilde{r} \,|\, \tilde{r} < VaR_\alpha(\tilde{r})\right\}.$$

2. Life Insurance Mathematics: Classical Approach

2.1. **Future Lifetime.** In the following the *future lifetime* or simply *lifetime* of a newborn person is described by a positive random variable $\tilde{X}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The cumulative distribution function of $\tilde{X}$ is

$$F(x) := \mathbb{P}(\tilde{X} \leq x), \text{ for } x \geq 0. \tag{82}$$

Unit of time is a year and $\tilde{X}$ is assumed to have a continuous distribution, that is, there exists a density function f such that $f = F'$ almost everywhere and

$$F(x) = \int_0^x f(u)du, \text{ for } x \geq 0. \tag{83}$$

In this framework, the probability that a newborn person dies between ages x and z (where $x < z$) can thus be expressed as

$$F(z) - F(x) = \int_x^z f(u)du.$$

Suppose now that a person is alive at age x, which mathematically means that we restrict our attention to the set $\{\tilde{X} > x\}$. Then the probability that this person dies before or at age $z > x$ is

$$\mathbb{P}(\tilde{X} \leq z | \tilde{X} > x) = \frac{\mathbb{P}\left(\{\tilde{X} \leq z\} \cap \{\tilde{X} > x\}\right)}{\mathbb{P}\left(\{\tilde{X} > x\}\right)} = \frac{F(z) - F(x)}{1 - F(x)}. \tag{84}$$

Definition 2.1.1: The *remaining lifetime at age* $x \geq 0$

$$\tilde{T}_x := (\tilde{X} - x) | \{\tilde{X} > x\} \tag{85}$$

is a positive random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The cumulative distribution function of remaining lifetime at age x is denoted

$${}_tq_x := \mathbb{P}(\tilde{T}_x \leq t) = \mathbb{P}(\tilde{X} - x \leq t | \tilde{X} > x), \text{ for } t \geq 0. \tag{86}$$

The probabilities concerning the remaining lifetime $\tilde{T}_x$ are interpreted as conditional probabilities according to equation (86).

The cumulative distribution function of remaining lifetime $\tilde{T}_x$ can be expressed in terms of the cumulative distribution function $F(x)$ of the lifetime of a newborn person as follows:

$${}_tq_x = \frac{F(t + x) - F(x)}{1 - F(x)}. \tag{87}$$

In addition to ${}_tq_x$, the following notations are conventional in classical life insurance mathematics: the probability that a person aged x lives at least t years more,

$${}_tp_x := \mathbb{P}(\tilde{T}_x > t) = 1 -_t q_x; \tag{88}$$

the probability that a person aged x lives at least t years more but dies before age $x+t+u$,

$${}_{t|u}q_x := \mathbb{P}(t < \tilde{T}_x \leq t + u) =_{t+u} q_x -_t q_x; \tag{89}$$

furthermore, it is conventional to denote $q_x :=_1 q_x$ and $p_x :=_1 p_x$. By properties of conditional expectations, we have the following identities:

$$ {}_tp_x = \frac{{}_{x+t}p_0}{{}_xp_0} \text{ and } {}_{t|u}q_x =_t p_x \cdot_u q_{x+t}. \tag{90} $$

In practice, the future lifetime of a person depends on many other factors besides the age of that person, such as health, gender, sosio-economic status and life style. Some of these factors are easier to take into account in pricing and underwriting of life insurance than others. Impact of health on mortality is intuitively obvious; gender has been empirically found to have so pronounced an impact on the future lifetime that different mortality distributions are used for men and women in life insurance. On the other hand, using life style as a factor in pricing may be difficult due to classification and measurement problems.

2.2. **Mortality.** Consider the probability that a person aged x dies within time period Δt divided with the length of the time period:

$$ \frac{{}_{\Delta t}q_x}{\Delta t} = \frac{\mathbb{P}\left(\tilde{T}_x \le \Delta t\right)}{\Delta t} = \frac{F(x+\Delta t) - F(x)}{\Delta t\,(1-F(x))} \tag{91} $$

Definition 2.2.1: The limit

$$ \mu_x := \lim_{\Delta t \to 0+} \frac{{}_{\Delta t}q_x}{\Delta t} = \frac{F^{'}(x)}{1-F(x)} \tag{92} $$

is called the *force of mortality at age x*[42] or simply *mortality at age x*.

Mortality $\mu_x \ge 0$ is defined for all values of x such that F is differentiable at x and $F(x) < 1$. Since

$$ \frac{d}{dx}\ln\left(1-F(x)\right) = -\frac{F^{'}(x)}{1-F(x)}, $$

mortality has a representation as a logarithmic derivative given by

$$ \mu_x = -\frac{d\ln(1-F(x))}{dx}. \tag{93} $$

The function

$$ s(x) := 1 - F(x) =_x p_0 \tag{94} $$

is called the *survival function*[43]. Equations (92) and (93) can be written in terms of survival function as a linear differential equation

$$ s^{'}(x) := -\mu_x \cdot s(x). \tag{95} $$

This equation is similar to the differential equation for technical provision (equation (12)) and highlights the mathematical similarity between force of interest δ and force of mortality μ.

[42](force of) mortality = dödlighets(intensitet)=kuolevuus(intensiteetti)

[43]survival function = överlevelsefunktion = selviytymisfunktio

Integrating equation (93), we obtain

$$\int_0^x \mu_u du = -|_0^x \ln s(u) = -\ln s(x),$$

for $F(x) < 1$, since $\ln s(0) = \ln(1) = 0$. Hence survival function has the representation

$$s(x) = \exp\left(-\int_0^x \mu_t dt\right). \tag{96}$$

This representation yields the following results:

(1) the conditional probability that a person aged x lives after t years is

$$_tp_x = \frac{s(x+t)}{s(x)} = \exp\left(-\int_0^t \mu_{x+u} du\right);$$

(2) the density function of remaining lifetime $\tilde{T}_x$ is

$$\frac{d}{dt}\,{}_tq_x = -\frac{d}{dt}\,{}_tp_x = \mu_{x+t} \cdot_t p_x;$$

(3) the expected value of remaining lifetime $\tilde{T}_x$ is

$$\mathring{e}_x := \mathbb{E}(\tilde{T}_x) = \int_0^\infty t \cdot \mu_{x+t} \cdot_t p_x dt = \int_0^\infty {}_tp_x dt.$$

We prove these assertions next.

Proof of (1.): By definition of survival function, $s(x) = {}_tp_0$, and combining this with equation (90) yields ${}_tp_x = s(x+t)/s(x)$. Furthermore, by representation (96),

$$\frac{s(x+t)}{s(x)} = \frac{\exp\left(-\int_0^{x+t} \mu_u du\right)}{\exp\left(-\int_0^x \mu_u du\right)} = \exp\left(-\int_x^{x+t} \mu_u du\right) = \exp\left(-\int_0^t \mu_{x+u} du\right). \square$$

Proof of (2.): Since ${}_tq_x = 1 - {}_tp_x$, property 1, which we just proved, implies that

$$\frac{d}{dt}\,{}_tq_x = -\frac{d}{dt}\,{}_tp_x = -\frac{d}{dt}\exp\left(-\int_0^t \mu_{x+u} du\right) = \mu_{x+t} \cdot \exp\left(-\int_0^t \mu_{x+u} du\right). \square$$

Proof of (3.): The first integral representation follows from property 2 combined with the definition of mathematical expectation, while the second integral representation follows from applying the general formula

$$\mathbb{E}(\tilde{Y}) = \int_0^\infty (1 - G(u))\, du$$

(valid for a non-negative random variable $\tilde{Y}$ and its c.d.f. G) to future lifetime $\tilde{T}_x$ and observing that by equation (90) $\mathbb{P}(\tilde{T}_x > t) = 1 - F(t) = {}_tp_x$. $\square$

By property 2 above

$$\mathbb{P}(\tilde{T}_x < \infty) = \int_0^\infty {}_tp_x \cdot \mu_{x+t} dt = 1, \tag{97}$$

in accordance with the obvious fact that a person aged x will eventually die in any case. Furthermore, the differential ${}_tp_x \cdot \mu_{x+t}dt$ can be interpreted as the probability that a person aged x is alive at age $x+t$ and dies during infinitesimal time interval dt after age $x+t$. Thus the probability that a person aged x is alive at age $x+n$ has the integral representation

$$\mathbb{P}(\tilde{T}_x > n) = \int_n^\infty {}_tp_x \cdot \mu_{x+t} dt. \tag{98}$$

There is a one-to-one correspondence between mortality functions μ and c.d.f.'s $F = 1-s$ of future lifetimes which possess a density. In order to formulate this result more precisely, we define $M := \sup\{x \mid s(x) > 0\}$, the smallest value of $x \in [-\infty, \infty]$ such that $s(x) = 0$. In particular, if the future lifetime has no finite upper bound, $M = \infty$.

Theorem: Suppose that future lifetime $\tilde{T}_x$ has a density function. Then there exists a function μ such that $\mathbb{P}(\tilde{T}_x < t) = 1 - s(t)$, where s is the survival function given in (94). Conversely, suppose that function μ is such that

$$\int_0^x \mu_t dt < \infty, \;\text{ for } 0 \le x < M, \tag{99}$$

and

$$\int_0^x \mu_t dt \to \int_0^M \mu_t dt = \infty, \; x \uparrow M. \tag{100}$$

Then

$$G(t) := 1 - \exp\left(-\int_0^t \mu_u du\right), \; 0 \le t < M, \tag{101}$$

is a c.d.f. of a continuous distribution such that $G(0) = 0$ and $G(M) = 1$. Moreover, μ satisfies equation (92) almost everywhere (with $F = G$).

Proof: Suppose first that $F(x) = 1 - s(x)$ is the c.d.f. of future lifetime. Since $s(0) = 1$ and s is continuous, necessarily $M > 0$. By representation (96), we have

$$\int_0^x \mu_t dt < \infty, \;\text{ for } 0 \le x < M, \tag{102}$$

and

$$\int_0^x \mu_t dt \to \int_0^M \mu_t dt = \infty, \; x \uparrow M. \tag{103}$$

Conversely, since μ is integrable over each subinterval $[0, x] \subset [0, M)$, function G in equation (101) is continuous and $G(0) = 1 - e^0 = 0$. Moreover, by equation (100) we have that $G(t) \to 1$ as $t \to M$. Finally, if μ is piecewise continuous, then G is piecewise continuously differentiable, which implies that

$$G^{'}(t) = \mu_t \exp\left(-\int_0^t \mu_u du\right) = \mu_t\left(1 - G(t)\right) \Leftrightarrow \mu_t = \frac{G^{'}(t)}{1 - G(t)}$$

for almost all t. The general case follows from this by using a truncation argument (see for example [8] pp. 49–50 for details). □

2.2.1. *Mortality Models.* In this section we introduce briefly some well-known mathematical models for mortality in chronological order. Mortality models are used in forecasting future mortality for the purposes of pricing life and pension insurance products and calculation of technical provisions. Hence it is important that the model should be able to fit relatively well into empirically observed mortality and moreover take into account the future trends in mortality – the strong decreasing trend in mortality observed in the developed countries during last century, due to (among other things) advances in medicine and the rising standard of living, has most often been underestimated by mortality models. While very positive things in themselves, the unanticipatedly strong decrease of mortality and increase of longevity can cause technical provisions and insurance premiums to be too low in relation to benefits promised.

a: De Moivre (1724):

$$\begin{aligned} F(t) &= \tfrac{t}{86}, \quad \text{for } 0 \le t \le 86, \\ \mu_t &= \tfrac{1}{86-t}, \quad \text{for } 0 \le t < 86. \end{aligned} \tag{104}$$

Future lifetime has an upper limit of 86 years in this mortality model. In the model mortality increases with age, but otherwise the behavior of the model is not really compatible with observed mortality. Observe that empirical mortality increases with age only after the very early childhood because of infant mortality, which causes mortality to be higher during the first year of life than in the years immediately after the first.

b: Gompertz (1824):

$$\mu_t = b \cdot e^{f \cdot (t-g)}. \tag{105}$$

In the model mortality increases with age in proportion to the level of mortality at that age, which can be represented in differential form as

$$d\mu_t = f \mu_t dt. \tag{106}$$

This yields a first order linear differential equation with constant coefficients, whose solution is given by (105).

Gompertz type mortality is used in the Finnish statutory occupational pension insurance (TyEL/ArPL). Parameter values for year 2009 are $b = 0.00005 \cdot e^{-0.57}$, $f = 0.095$, and value of g differs depending on gender and age cohort (determined by during which decade the person concerned was born). Similar piecewise Gompertz force of mortality is also used in the reference mortality model of the Finnish statutory accident insurance.

c: Makeham (1860):

$$\mu_t = a + b \cdot e^{f \cdot (t-g)}. \tag{107}$$

This is similar to Gompertz model, except that there is an additional age-independent parameter a which describes the possibility of death due to accidents.

The reference mortality model used in Finnish non-statutory life insurance has a force of mortality which can be seen as a modification of the Makeham model.

d: Weibull (1939):

$$\mu_t = bt^n. \tag{108}$$

e: Logistic mortality:

$$\mu_t = \frac{b \cdot e^t}{1 + d \cdot e^t}. \tag{109}$$

In the logistic model, mortality increases more gradually with t for high values of t than in Makeham model. With suitably chosen parameter values logistic mortality fits better to observed mortality for advanced ages than Makeham type mortality.

2.2.2. *Select Mortality and Cohort Mortality.* A life insurer often requires a medical examination of the health of the potential insured person prior to underwriting a life policy. For this reason newly insured people are on average more healthy than people of same age who have been insured for a longer time. Moreover, the mortality of the insured parties is usually lower than that of the population as a whole.

In *select mortality*, the mortality of the insured persons depends also on how long they have been insured. The effect of initial selection diminishes as time passes. In select mortality, we denote with ${}_tq_{[x-v]+v}$ the c.d.f. of the future lifetime $\tilde{T}_x$ of a person aged x who has been insured from age $x - v$. For an insured person aged x,

$${}_tq_x < {}_tq_{[x-1]+1} < \ldots < {}_tq_{[x-r]+r} = {}_tq_{[x-s]+s}, \tag{110}$$

for all $s > r$, where r is the time period during which the effect of selection vanishes.

Similar to select mortality is *cohort mortality.* Due to advances in medicine, improved living standards and other such reasons, the expected future lifetime of new generations may be longer than that of earlier generations. Long-lasting development in this direction manifests itself as a decreasing trend in mortality.

2.2.3. *Competing Causes of Death.* In the theory of competing causes of death a central objective is to find out how removing a specific cause of death impacts overall mortality. To construct the theory, a basic assumption is made that each death can be associated with a unique cause of death, which is a rather stringent assumption.

We assume that for each death, the age of the deceased and the cause of death are known. Based on these we can estimate the probability of dying of cause k by age t:

$$G_k(t) := \mathbb{P}\left(\tilde{X} \le t \text{ and cause of death is } k\right), \tag{111}$$

where $\tilde{X}$ is the lifetime of a (newborn) individual. Function G_k is not a c.d.f., since

$$\lim_{t\to\infty} G_k(t) = \mathbb{P}(\text{ cause of death is } k) < 1.$$

For c.d.f. F of lifetime $\tilde{X}$ we have

$$F = \sum_k G_k, \tag{112}$$

since each death was assumed to be associated with a single, uniquely determined cause k.

In addition to probabilities $G_k(t)$, we denote by $F_k(t)$ the probability that an individual dies of cause k by age t, *assuming that k is the only possible cause of death.* This probability is hypothetical and cannot be directly observed or estimated.

Precisely stated, lifetime $\tilde{X}$ is assumed to satisfy

$$\tilde{X} = \min_k \tilde{X}_k, \tag{113}$$

where $\tilde{X}_k$ is lifetime assuming that k is the only possible cause of death. This representation explains the terminology: different causes of death "compete" on which of them first succeeds in causing the death of an individual. By many different means (e.g. development of medicine) the society attempts to remove or at least mitigate causes of death, which leads to increases in lifetime $\tilde{X}$.

If causes of death are assumed to be mutually independent, the survival function can be expressed as a product of survival functions related to individual causes of death:

$$s(t) = 1-F(t) = \mathbb{P}(\tilde{X} > t) = \mathbb{P}\left(\cap_k\{\tilde{X}_k > t\}\right) = \prod_k \mathbb{P}(\tilde{X}_k > t) = \prod_k (1 - F_k(t)) = \prod_k s_k(t). \tag{114}$$

In contrast, equation (112) does not require independence, but only that at most one cause of death can be realized with positive probability simultaneously.

An important problem is what happens to probabilities $F(t)$ and $G_k(t)$, if some cause of death (say h) is removed. Let us denote the probabilities after removing cause h by $F_{\circ h}(t)$ and $G_{k\circ h}(t)$, for $k \neq h$. We will next establish that if causes of death are mutually independent, then these probabilities can be calculated provided that G_k and F_k are known.

Let μ be the force of mortality corresponding to the c.d.f. of lifetime $\tilde{X}$ and let μ_k be the force of mortality corresponding to the c.d.f. of hypothetical lifetime $\tilde{X}_k$ (i.e. μ_k is the mortality associated with cause k in the case that k is the only possible cause of death). Then

$$\mu = \frac{F'}{1-F} \text{ and } \mu_k = \frac{F_k'}{1-F_k}. \tag{115}$$

Let also ν_k be the force of mortality associated with cause of death k in the case all causes of death are possible. This mortality is defined as the limit

$$\begin{aligned} \nu_k(t) &:= \lim_{\Delta t\to 0+} \frac{\mathbb{P}(t<\tilde{X}\le t+\Delta t \text{ and } \tilde{X}=\tilde{X}_k \mid \tilde{X}>t)}{\Delta t} \\ &= \lim_{\Delta t\to 0+} \frac{G_k(t+\Delta t)-G_k(t)}{\Delta t(1-F(t))} = \frac{G_k'(t)}{1-F(t)}. \end{aligned} \tag{116}$$

Total mortality corresponding to lifetime $\tilde{X}$ and its c.d.f. F is the sum of mortalities ν_k:

$$\mu = \sum_k \nu_k. \tag{117}$$

If causes of death are mutually independent, i.e. if random variables $\tilde{X}_k$ are mutually independent, then

$$\nu_k = \mu_k, \text{ for each } k. \tag{118}$$

To see this, observe that by definitions of G_k and lifetime $\tilde{X}$, $G_k(t) = \mathbb{P}(X \leq t \text{ and } \tilde{X} = \tilde{X}_k)$. By virtue of mutual independence of variables $\tilde{X}_k$ we can apply the formula for total probability to get

$$\begin{aligned} G_k(t) &= \textstyle\int_0^\infty \mathbb{P}\left(\tilde{X} \leq t \text{ and } \tilde{X} = \tilde{X}_k \,|\, \tilde{X}_k = y\right) dF_k(y) \\ &= \textstyle\int_0^t \mathbb{P}\left(\tilde{X} = \tilde{X}_k \,|\, \tilde{X}_k = y\right) dF_k(y) = \int_0^t \mathbb{P}\left(\cap_{j\neq k}\{\tilde{X}_k > y\}\right) dF_k(y) \\ &= \textstyle\int_0^t \prod_{j\neq k} \mathbb{P}\left(\{\tilde{X}_k > y\}\right) dF_k(y) = \int_0^t \prod_{j\neq k}(1 - F_j(y)) dF_k(y) \\ &= \textstyle\int_0^t F_k^{'}(y) \prod_{j\neq k}(1 - F_j(y)) dy \end{aligned}$$

Differentiation yields $G_k^{'} = F_k^{'} \prod_{j\neq k}(1 - F_j)$. On the other hand, by independence and (114) $1 - F = \prod_k (1 - F_k)$, so that

$$\nu_k = \frac{G_k^{'}}{1-F} = F_k^{'} \frac{\prod_{j\neq k}(1 - F_j)}{\prod_j (1 - F_j} = \frac{F_k^{'}}{1 - F_k} = \mu_k.$$

Intensities μ_k and ν_k are equal, even though functions G_k and F_k are different. From this we immediately get that

$$\mu_k = \sum_k \mu_k \text{ and } \mu_{\circ h} = \sum_{k\neq h} \mu_k = \mu - \mu_h,$$

where $\mu_{\circ h}$ is the mortality associated with probability distribution $F_{\circ h}$.

Probability distributions F, F_k and $F_{\circ h}$ have the usual integral representations in terms of mortalities μ, μ_k and $\mu_{\circ h} = \mu - \mu_h$. We conclude this section by deriving similar integral representations for G_k and $G_{k\circ h}$. Since $G_k^{'} = \mu_k(1 - F)$, by equation (96)

$$G_k(t) = \int_0^t G_k^{'}(y)dy = \int_0^t \mu_k(y)(1 - F(y))dy = \int_0^t \mu_k(y) \exp\left(-\int_0^y \mu(s)ds\right) dy. \tag{119}$$

By applying this to the case where cause of death h has been removed, we get

$$G_{k\circ h}(t) = \int_0^t \mu_k(y) \exp\left(-\int_0^y (\mu(s) - \mu_h(s))ds\right) dy. \tag{120}$$

2.3. **Expected Present Values of Life Insurance Contracts.** In this chapter we derive the expected present value of the insurer's payment stream for some fundamental life and pension insurance contracts. These expected present values are called *net single premiums.* Net premium covers the insurer's expected claim costs but not its operating expenses.

2.3.1. *Mortality and Interest.* Suppose that an insurer has made a contract stipulating that it pays a person aged x the sum S, if this person is alive after time n, and otherwise nothing. Given a force of interest δ, the (random) present value A of this payment by the insurer discounted to beginning of contract period is

$$A = \begin{cases} Se^{-\delta n}, & \text{if person is alive after time } n \\ 0, & \text{otherwise,} \end{cases} \tag{121}$$

and taking expectations yields

$$\mathbb{E}(A) = S \cdot e^{-\delta n} \cdot {}_n\, p_x = S \cdot e^{-\delta n} \cdot \frac{{}_{x+n}p_0}{{}_xp_0}. \tag{122}$$

Denote now

$$D_x := e^{-\delta x} \cdot {}_x\, p_0 = v^x \cdot {}_x\, p_0 \tag{123}$$

to see that the expected value of the payment stream (121), discounted to initial time, is

$$\mathbb{E}(A) = S \cdot \frac{D_{x+n}}{D_x}. \tag{124}$$

Quotient D_{x+n}/D_x is the mortality-dependent discount factor for *life annuities*, which takes into account the impacts of both interest rate and mortality when calculating present values. Formally, mortality and interest rate influence this discount factor completely analogously, as is apparent from the following formulas:

$$\frac{D_{x+t}}{D_x} = \exp\left(-\int_x^{x+t} (\delta + \mu_u) du\right) \tag{125}$$

and

$$D_x' = -D_x \cdot (\delta + \mu_x). \tag{126}$$

These formulas are easily proved by recalling that (96) implies that ${}_xp_0 = \exp\left(-\int_0^x \mu_s ds\right)$ and hence by virtue of (123),

$$D_x = \exp\left(-\int_0^x (\delta + \mu_s) ds\right),$$

which implies (125) and (126). In particular, if mortality equals zero on $(x, x+t)$, then $D_{x+t}/D_x = v^t$, the usual discount factor.

Example: Suppose that effective annual interest rate is 10 % and mortality is a constant such that the probability of death within a year is 1 %. Then force of interest $\delta = \ln(1 + 0,10) \approx 0,0953$ and mortality $\mu = -\ln(0,99) \approx 0,01005$. Hence $D_{x+1}/D_x = \exp(\ln(0,99) - \ln(1,10)) = 0,99/1,10 = 0,9$. The joint impact of interest and mortality on the present value is thus equivalent to a discount rate $1/0,9 - 1 \approx 11,1$ %.

Definition 2.3.1: Function D_x and functions

$$\overline{N}_x := \int_x^\infty D_t dt \tag{127}$$

and

$$\overline{M}_x := \int_x^\infty D_t \mu_t dt \tag{128}$$

defined in terms of it are called *commutation functions.*

Commutation functions are finite-valued, if $\delta > 0$ or, more generally, $\delta + \mu_x \geq c > 0$ for x sufficiently large. We assume that at least the latter condition is satisfied, as is the case in almost all practical cases. The overbar in notations $\overline{N}_x$ and $\overline{M}_x$ is a standard convention in life insurance mathematics; notations N_x and M_x refer to the corresponding commutation functions in the discrete time model.

Commutation functions are handy in expressing and tabulating the joint impact of interest and mortality in life and pension insurance calculations.

The following relation holds between commutation functions:

$$\overline{M}_x = D_x - \delta \overline{N}_x. \tag{129}$$

To see this, observe that $D_\infty = 0$ and hence by (126)

$$D_x = -\int_x^\infty D_t' dt = \delta \int_x^\infty D_t dt + \int_x^\infty D_t \mu_t dt = \delta \overline{N}_x + \overline{M}_x.$$

2.3.2. *Present Value of A Single Life Insurance Contract.* In life insurance the *sum insured* is paid to the *beneficiaries* upon death or survival of the insured person.

Consider a life insurance contract where the insured person is currently at age x. The insurance contract terminates after the pre-specified *term* n of the contract at termination age $w = x + n$. It is possible to have $n = \infty$ (in which case the policy is in force for the whole lifetime of the insured person).

Definition 2.3.2: **i:** In a *pure endowment insurance*[44] (notation: **V**/ S / n) the sum insured S is paid to the beneficiary at the end of the contract term, provided that the insured person attains termination age $w = x + n$. In case the insured dies before the end of the term, nothing is paid to the beneficiary.

ii: In a *term life insurance*[45] (notation: **K**/ S / n) the sum insured S is paid to the beneficiary upon the death of the insured person, provided that this happens before termination age $w = x + n$. In case the insured is alive at the end of the term, nothing is paid to the beneficiary.

[44] pure endowment insurance = kapitalförsäkring för livsfall = elämänvaravakuutus

[45] term life insurance = kapitalförsäkring för dödsfall = kuolemanvaravakuutus

iii: A *whole life insurance*[46](notation: **K**/ S / ∞) is a term life insurance with infinite term ($n = \infty$).

iv: In an *endowment life insurance*[47] (notation: **Y**/ S / n)the sum insured S is paid to the beneficiary either upon the death of the insured person or at the termination age $w = x + n$, whichever occurs first.

Endowment life insurance is, of course, a combination of a term life and a pure endowment insurance.

Next we calculate the *random present value* $\tilde{A}_{x:\overline{n}|}$ of a unit sum insured ($S = 1$) potentially paid to the beneficiary for different insurance contracts.

For a pure endowment insurance contract

$$\tilde{A}_{x:\overline{n}|}(\mathbf{V}) = \begin{cases} e^{-\delta \cdot n} & , \text{ if } \tilde{T}_x > n \\ 0 & , \text{ if } \tilde{T}_x \leq n, \end{cases} \tag{130}$$

since the unit payment is made after time n has passed, provided that the insured person is alive then (i.e. $\tilde{T}_x > n$).

For a term life insurance contract

$$\tilde{A}_{x:\overline{n}|}(\mathbf{K}) = \begin{cases} 0 & , \text{ if } \tilde{T}_x > n \\ e^{-\delta \cdot \tilde{T}_x} & , \text{ if } \tilde{T}_x \leq n, \end{cases} \tag{131}$$

since the unit payment is made upon the death of the insured after time $\tilde{T}_x$ has passed, provided that the contract term has not expired (i.e. $\tilde{T}_x \leq n$).

For an endowment life insurance contract

$$\tilde{A}_{x:\overline{n}|}(\mathbf{Y}) = \begin{cases} e^{-\delta \cdot n} & , \text{ if } \tilde{T}_x > n \\ e^{-\delta \cdot \tilde{T}_x} & , \text{ if } \tilde{T}_x \leq n, \end{cases} \tag{132}$$

since the unit payment is made upon the death of the insured after time $\tilde{T}_x$ has passed, or at the end of the contract term after time n has passed, whichever occurs first. In particular we observe that $\tilde{A}_{x:\overline{n}|}(\mathbf{Y}) = \tilde{A}_{x:\overline{n}|}(\mathbf{V}) + \tilde{A}_{x:\overline{n}|}(\mathbf{K})$.

These random present values can be represented using indicator functions as follows:

$$\tilde{A}_{x:\overline{n}|}(\mathbf{V}) = I_{(n,\infty)}(\tilde{T}_x) \cdot e^{-\delta n}$$

$$\tilde{A}_{x:\overline{n}|}(\mathbf{K}) = I_{(0,n]}(\tilde{T}_x) \cdot e^{-\delta \tilde{T}_x}$$

$$\tilde{A}_{x:\overline{n}|}(\mathbf{Y}) = \tilde{A}_{x:\overline{n}|}(\mathbf{V}) + \tilde{A}_{x:\overline{n}|}(\mathbf{K}) = e^{-\delta \cdot \min(n, \tilde{T}_x)}.$$

For a whole life insurance contract, we use notation

$$\tilde{A}_x(\mathbf{K}) := \tilde{A}_{x:\overline{\infty}|}(\mathbf{K}).$$

[46]whole life insurance = oändlig kapitalförsäkring för dödsfall = elinikäinen kuolemanvaravakuutus

[47]endowment life insurance = sammansatt kapitalförsäkring = yhdistetty vakuutus

The previous present values are random variables, because they depend on the remaining lifetime of the insured person, which is random. For the purpose of pricing insurance contracts, we are interested in the expected values of these random present values

$$A_{x:\overline{n}|}(\cdot) := \mathbb{E}\left[\tilde{A}_{x:\overline{n}|}(\cdot)\right],$$

which are called *net single premiums*[48] of the corresponding insurance contracts.

Theorem: The expected value $A_{x:\overline{n}|}(\cdot)$ of the random present value of a unit sum insured $S = 1$ at the time when the insured person is at age x, is

i: for pure endowment insurance

$$A_{x:\overline{n}|}(\mathbf{V}) = e^{-\delta n} \cdot {}_{n}p_x = \frac{D_{x+n}}{D_x};$$

ii: for term life insurance

$$A_{x:\overline{n}|}(\mathbf{K}) = \int_0^n \frac{D_{x+t}}{D_x} \cdot \mu_{x+t} dt = \frac{\overline{M}_x - \overline{M}_{x+n}}{D_x};$$

iii: for endowment life insurance

$$A_{x:\overline{n}|}(\mathbf{Y}) = A_{x:\overline{n}|}(\mathbf{V}) + A_{x:\overline{n}|}(\mathbf{K}).$$

Proof: (i): This was proved previously, see Equation (124).

(ii): Denote $F_x(t) := {}_{t}q_x = 1 - {}_{t}p_x$. Then

$$\begin{aligned} A_{x:\overline{n}|}(\mathbf{K}) &= \mathbb{E}\left[e^{-\delta \tilde{T}_x} \cdot I_{(0,n]}(\tilde{T}_x)\right] = \int_0^n e^{-\delta t} dF_x(t) = \int_0^n e^{-\delta t} \cdot \mu_{x+t} \cdot {}_{t}p_x dt \\ &= \int_x^{x+n} e^{-\delta(s-x)} \cdot \mu_s \cdot \frac{{}_{s}p_0}{{}_{x}p_0} ds = \int_x^{x+n} \frac{D_s}{D_x} \cdot \mu_s ds = \frac{\overline{M}_x - \overline{M}_{x+n}}{D_x}. \square \end{aligned}$$

By virtue of the representations of the expected present values in terms of indicator functions given above, we can establish that the variance of the random present value

$$Var\left(S \cdot \tilde{A}_{x:\overline{n}|}(\cdot)\right) = S^2 \cdot \left[A^{(2\delta)}_{x:\overline{n}|}(\cdot) - \left(A_{x:\overline{n}|}(\cdot)\right)^2\right]$$

for each insurance type ($\mathbf{V}$, $\mathbf{K}$ and $\mathbf{Y}$) considered, where $A^{(2\delta)}_{x:\overline{n}|}(\cdot)$ is the expected present value of the contract computed with twofold force of interest 2δ. This follows from the fact that for indicator functions it is true that $(I_B(t))^2 = I_B(t)$ (indicator functions are *idempotent*).

2.3.3. *Present Value of A Pension.* Suppose that for a certain pre-specified period of time, pension payments of annual magnitude E are paid continuously to a person while that person is alive. Such arrangement is called a *continuous life annuity.* Continuous payments mean that pension payments accrued in an infinitesimal time interval dt equal $E \cdot dt$. The difference to continuous annuities considered previously in this presentation is that payments are conditional on the beneficiary being alive.

[48] net single premium = engångsbetalning = nettokertamaksu

Definition 2.3.3: Suppose that a pension of annual magnitude $E = 1$ is continuously paid to a person currently aged x for as long as this person is alive. The expected present value of this *continuous whole life unit annuity* for a person aged x is denoted by

$$\overline{a}_x := \mathbb{E}(\overline{a}_{\overline{\tilde{T}_x|}}). \tag{133}$$

If the annuity is such that the pension is paid at most for n years after age x, then it is called a *continuous n-year unit life annuity*, and its expected present value is

$$\overline{a}_{x:\overline{n}|} := \mathbb{E}(\overline{a}_{\overline{\min(n,\tilde{T}_x)|}}). \tag{134}$$

The expected present value of a n-year unit pension can be represented in terms of the indicator function of a person aged x being alive after t years,

$$\tilde{J}_x(t) := I_{(t,\infty)}(\tilde{T}_x), \tag{135}$$

as

$$\overline{a}_{x:\overline{n}|} = \mathbb{E}\left(\int_0^n \tilde{J}_x(t) \cdot e^{\delta t} dt\right) = \int_0^n \mathbb{E}\left(\tilde{J}_x(t)\right) \cdot e^{\delta t} dt. \tag{136}$$

Theorem: The expected present values of continuous life annuities are given by the following expressions:

i: $\overline{a}_x = \int_0^\infty {}_tp_x \cdot e^{-\delta t} dt = \int_0^\infty \frac{D_{x+t}}{D_x} dt = \frac{\overline{N}_x}{D_x}$;

ii: $\overline{a}_{x:\overline{n}|} = \int_0^n {}_tp_x \cdot e^{-\delta t} dt = \int_0^n \frac{D_{x+t}}{D_x} dt = \frac{\overline{N}_x - \overline{N}_{x+n}}{D_x}$;

iii: $\overline{a}_{x:\overline{n}|} = \overline{a}_x - \frac{D_{x+n}}{D_x} \cdot \overline{a}_{x+n}$.

Proof: (ii) follows from Equation (136), since

$$\mathbb{E}\left(\tilde{J}_x(t)\right) = \mathbb{E}\left(I_{(t,\infty)}(\tilde{T}_x)\right) = \mathbb{P}\left(\tilde{T}_x > t\right) = {}_tp_x.$$

(i) is a special case of (ii), as $\overline{N}_\infty = 0$.

(iii) $\overline{a}_{x:\overline{n}|} = \frac{\overline{N}_x}{D_x} - \frac{\overline{N}_{x+n}}{D_{x+n}} \frac{D_{x+n}}{D_x} = \overline{a}_x - \frac{D_{x+n}}{D_x} \overline{a}_{x+n}$. □

Definition 2.3.4: In a *continuous pension insurance beginning at age x* (notation: **E** / E / m / n) the insured begins to receive pension payments of annual magnitude E after time $m \geq 0$ at age $w = x + m$. The pension payments are made as long as the insured is alive, however for at most n years. Such a pension arrangement is called *m years delayed n-year continuous life annuity*; if $n = \infty$, then we speak of a *whole life* annuity.

In a pension insurance contract the payments of the life annuity begin at age $x + m$, and the duration of the life annuity is a random variable depending on the remaining lifetime of a person aged x:

$$\tilde{Y} = \min\left(n, \max(0, \tilde{T}_x - m)\right) = \begin{cases} 0 & , \text{ if } \tilde{T}_x \leq m \\ \tilde{T}_x - m & , \text{ if } m < \tilde{T}_x \leq m + n \\ n & , \text{ if } \tilde{T}_x \geq m + n; \end{cases}$$

in particular, should the insured person die before age $x+m$, no pension payments are made. Hence the *expected present value of an m years delayed n-year continuous unit life annuity* at the time when the insured is aged x is

$$ {}_{m|}\overline{a}_{x:\overline{n}|} := \mathbb{E}\left(v^m \cdot \overline{a}_{\overline{\tilde{Y}}|}\right), \tag{137} $$

where $v = e^{-\delta}$ is the discount factor. As a special case we get the expected present value of an m years delayed continuous unit whole life annuity

$$ {}_{m|}\overline{a}_x := \mathbb{E}\left(v^m \cdot \overline{a}_{\overline{\max(0,\tilde{T}_x - m)}|}\right). \tag{138} $$

We have the following relations between the expected present values of delayed life annuities and "undelayed" annuities:

i: ${}_{m|}\overline{a}_{x:\overline{n}|} = \frac{D_{x+m}}{D_x} \cdot \overline{a}_{x+m:\overline{n}|}$;
ii: ${}_{m|}\overline{a}_x = \frac{D_{x+m}}{D_x} \cdot \overline{a}_{x+m}$;
iii: $\overline{a}_{x:\overline{n}|} = \overline{a}_x - {}_{n|}\,\overline{a}_x$.

To see this, observe that given $\tilde{T}_x > m$ we have $\tilde{T}_x - m = \tilde{T}_{x+m}$, that is, $\tilde{Y} = \min(n, \tilde{T}_{x+m})$. Then

$$ \mathbb{E}\left(\overline{a}_{\overline{\tilde{Y}}|}\right) = {}_m q_x \cdot \mathbb{E}\left(\overline{a}_{\overline{0}|}\right) + {}_m p_x \cdot \mathbb{E}\left(\overline{a}_{\overline{\min(n,\tilde{T}_{x+m})}|}\right) = {}_m p_x \cdot \overline{a}_{x+m:\overline{n}|}, $$

since ${}_m q_x = \mathbb{P}(\tilde{T}_x \le m)$, ${}_m p_x = \mathbb{P}(\tilde{T}_x > m)$, and $\tilde{Y} = 0$ in case $\tilde{T}_x \le m$. Formula (i) follows by discounting the above expected value to the moment of time m years earlier with discount factor v^m and recalling that

$$ v^m \cdot {}_m p_x = \frac{D_{x+m}}{D_x}. $$

Formula (ii) is a special case of (i) (for $n \to \infty$), and (iii) follows from part (iii) of previous theorem.

Similarly to formula (iii) above, we have the following relationship between the expected present values of term life and whole life insurance contracts :

$$ A_{x:\overline{n}|}(\mathbf{K}) = A_x(\mathbf{K}) - {}_{n|}\,A_x(\mathbf{K}) $$

where

$$ {}_{n|}A_x(\mathbf{K}) := \frac{\overline{M}_{x+n}}{D_x} = \frac{D_{x+n}}{D_x} \cdot \frac{\overline{M}_{x+n}}{D_{x+n}} = \frac{D_{x+n}}{D_x} \cdot A_{x+n}(\mathbf{K}). $$

The expected present values of life annuities can be represented in terms of the expected present values of life insurance contracts as follows:

$$ \overline{a}_x = \frac{1 - A_x(\mathbf{K})}{\delta} \text{ and } \overline{a}_{x:\overline{n}|} = \frac{1 - A_{x:\overline{n}|}(\mathbf{Y})}{\delta}. $$

The first result is easily shown using previous results:

$$ \delta \cdot \overline{a}_x = \delta \cdot \frac{\overline{N}_x}{D_x} = 1 - \frac{\overline{M}_x}{D_x} = 1 - A_x(\mathbf{K}) $$

where we have used part (i) of previous Theorem, Equation (129) and part (ii) of the Theorem before the previous Theorem. The second result is obtained using additionally part (iii) of the Theorem before the previous Theorem:

$$\begin{aligned}
\delta \cdot \overline{a}_{x:\overline{n}|} &= \delta \cdot \overline{a}_x - \tfrac{D_{x+n}}{D_x} \cdot \delta \cdot \overline{a}_{x+n} \\
&= 1 - \tfrac{\overline{M}_x}{D_x} - \tfrac{D_{x+n}}{D_x} + \tfrac{D_{x+n}}{D_x} \cdot \tfrac{\overline{M}_{x+n}}{D_{x+n}} \\
&= 1 - \tfrac{D_{x+n}}{D_x} - \tfrac{\overline{M}_x - \overline{M}_{x+n}}{D_x} \\
&= 1 - A_{x:\overline{n}|}(\mathbf{V}) - A_{x:\overline{n}|}(\mathbf{K}) = 1 - A_{x:\overline{n}|}(\mathbf{Y}).
\end{aligned}$$

The formulas' interpretation can be simplified by rearranging them; part (i), for example, can be written as

$$1 = \delta \cdot \overline{a}_x + A_x(\mathbf{K}),$$

which can be interpreted as follows: with a unit payment, the insured can obtain continuous interest payments with force of interest δ for as long as she lives and a unit sum insured payable upon the death of the insured.

In practice, pensions are not paid continuously but at regular intervals, e.g. monthly. Next we look at the discrete analogues of the expected present values of continuous life annuities.

Definition 2.3.5: Let m and $m \cdot n$ be positive integers (so that n is a multiple of fraction $1/m$). Then the expected present value of an *immediate whole life unit annuity payable m times a year* to a person aged x is

$$a_x^{(m)} := \mathbb{E}\left(a^{(m)}_{\overline{\tilde{T}_x}|}\right),$$

and the expected present value of an *immediate n-year unit life annuity payable m times a year* is

$$a^{(m)}_{x:\overline{n}|} := \mathbb{E}\left(a^{(m)}_{\overline{\min(n,\tilde{T}_x)}|}\right).$$

Corresponding expected present values for unit life annuities due are

$$\ddot{a}_x^{(m)} := \mathbb{E}\left(\ddot{a}^{(m)}_{\overline{\tilde{T}_x}|}\right) \text{ and } \ddot{a}^{(m)}_{x:\overline{n}|} := \mathbb{E}\left(\ddot{a}^{(m)}_{\overline{\min(n,\tilde{T}_x)}|}\right).$$

In case $m = 1$, the superscript is usually omitted.

Remaining lifetime $\tilde{T}_x$ has a continuous distribution, and hence it can take on any values. Incomplete periods are disregarded when calculating the expected values of discrete life annuities.

Following formulas hold for discrete life annuities:

i: $a^{(m)}_{x:\overline{n}|} = \frac{1}{m}\sum_{k=1}^{nm} \frac{D_{x+k/m}}{D_x}$;

ii: $\ddot{a}^{(m)}_{x:\overline{n}|} = \frac{1}{m}\sum_{k=0}^{nm-1} \frac{D_{x+k/m}}{D_x}$;

iii: $a_x^{(m)} = \frac{1}{m}\sum_{k=1}^{\infty}\frac{D_{x+k/m}}{D_x}$;
iv: $a_{x:\overline{n}|}^{(m)} = \frac{D_{x+1/m}}{D_x}\cdot \ddot{a}_{x:\overline{n}|}^{(m)}$;
v: $\ddot{a}_{x:\overline{n}|}^{(m)} = \frac{1}{m} + a_{\overline{x:n-1/m}|}^{(m)}$.

We prove property (i): in a m times a year payable immediate life annuity, the kth payment is made if the beneficiary is alive at time k/m when the payment is due. In other words, if the indicator of being alive $\tilde{J}_x\left(\frac{k}{m}\right)$ takes the value 1. Since $\mathbb{E}\left[\tilde{J}_x\left(\frac{k}{m}\right)\right] = {}_{k/m}p_x$, we get

$$\begin{aligned} a_{x:\overline{n}|}^{(m)} &= \mathbb{E}\left(a_{\overline{\min(n,\tilde{T}_x)}}^{(m)}\right) = \mathbb{E}\left(\frac{1}{m}\sum_{k=1}^{nm} e^{-\delta\cdot k/m}\cdot \tilde{J}_x\left(\frac{k}{m}\right)\right) \\ &= \frac{1}{m}\sum_{k=1}^{nm} e^{-\delta\cdot k/m}\cdot {}_{k/m}p_x = \frac{1}{m}\cdot\sum_{k=1}^{nm}\frac{D_{x+k/m}}{D_x}. \end{aligned}$$

Introducing the discrete commutation function

$$N_x^{(m)} := \frac{1}{m}\cdot\sum_{k=0}^{\infty} D_{x+k/m} \tag{139}$$

we obtain the following analogues for discrete life annuities of previous formulas for continuous life annuities:

$$\begin{aligned} a_x^{(m)} &= \frac{N_{x+1/m}^{(m)}}{D_x}, & \ddot{a}_x^{(m)} &= \frac{N_x^{(m)}}{D_x}, \\ a_{x:\overline{n}|}^{(m)} &= \frac{N_{x+1/m}^{(m)} - N_{x+n+1/m}^{(m)}}{D_x}, & \ddot{a}_{x:\overline{n}|}^{(m)} &= \frac{N_x^{(m)} - N_{x+n}^{(m)}}{D_x}, \\ a_{x:\overline{n}|}^{(m)} &= a_x^{(m)} - \frac{D_{x+n}}{D_x}\cdot a_{x+n}^{(m)}, & \ddot{a}_{x:\overline{n}|}^{(m)} &= \ddot{a}_x^{(m)} - \frac{D_{x+n}}{D_x}\cdot \ddot{a}_{x+n}^{(m)}. \end{aligned}$$

2.3.4. *Net Premiums.* In the previous Section we calculated for different insurance contracts the *net single premium*, that is, the expected present value of future claim or pension payments at the beginning of the contract. "Net" refers to the fact that the premium consists of expected future claim costs only and does not incorporate any additional expense loadings which would cover claim handling expenses or other similar expenses that the insurer incurs.

In practice, premiums are usually not paid as a single advance payment but instead as periodic payments.

Many life and pension insurance products are essentially saving instruments: for example in a pure endowment insurance the premiums are in a way deposited into the insurance company which pays back to the insured the deposited sum with interest at the maturity of the contract. In such products the delay between payment of premiums and of claims is typically very long.

In contrast, products such as term life which are taken to obtain coverage for specific risks, the delay is typically much shorter. In principle, the net premium of a term life insurance contract could be paid as a so-called *natural premium* which equals the claim intensity $S \cdot \mu_t$. In this case the premium for an infinitesimal time period $[t, t + dt]$ equals the expected value of claim costs for the time period, $S \cdot \mu_t dt$. This premium changes with the age of the insured and it is paid for as long as the insured is alive. Applied interest rate has no effect on the premium, because (in expected value sense) there is no delay between premiums and claim payments.

Definition 2.3.6: Let the age of the insured at the beginning of the contract be x and suppose that the premium is paid until age $x + h$ for as long as the insured is alive. Then h is the *premium payment period* of the insurance contract.

Continuous *net level premium* $\overline{P}$ is the constant continuous premium whose annual amount is $\overline{P}$ and which is paid during the premium payment period for as long as the insured is alive. Premium during time interval $[t, t + dt]$ is $\overline{P}dt$, provided that the insured is alive.

Discrete *m times a year payable net premium due* $\ddot{P}^{(m)}$ is paid during the premium payment period for as long as the insured is alive in equally large installments periodically so that period length is constant and total annual payment equals $\ddot{P}^{(m)}$.

There are many different methods to collecting premiums. In comparisons between alternatives, the impact of interest needs to be taken into account.

Definition 2.3.7: Consider (possibly random) cash flows.

- **i:** Two cash flows (under the same interest rate assumptions) are *equivalent*, if their expected present values are equal;
- **ii:** If two cash flows have different interest rate assumptions, their expected present values can be equal at a specified moment of time t. Then these cash flows are under the assumed interest rates *equivalent at time* t;
- **iii:** *Equivalence principle* refers to adjusting (possibly random) cash flows to be equivalent (at a specified moment of time).

Assumed forces of interest need not be constants.

Equivalence principle is usually applied when setting premium levels in life or pension insurance to match the expected claim costs and other expenses.

In the following we assume that the insured is aged x at the beginning of the insurance contract and that the premium payment period is h; additionally, force of interest δ is assumed to be constant.

The continuous level premium $\overline{P}$ equivalent to the net single premium A can be obtained in accordance with equivalence principle from equation

$$\overline{P} \cdot \overline{a}_{x:\overline{h}|} = A. \tag{140}$$

In similar fashion, the discrete m times a year payable net premium due $\ddot{P}^{(m)}$ equivalent to the net single premium A can be obtained in accordance with equivalence principle from equation

$$\ddot{P}^{(m)} \cdot \ddot{a}^{(m)}_{x:\overline{h}|} = A. \tag{141}$$

Example: For a term life insurance $(\mathbf{K}/S/n)$ with premium payment period n, equation (141) takes the form

$$\overline{P} \cdot \overline{a}_{x:\overline{n}|} = S \cdot A_{x:\overline{n}|}(\mathbf{K}).$$

For a pension insurance $(\mathbf{E}/E/h/\infty)$ with premium payment period h, equation (141) takes the form

$$\overline{P} \cdot \overline{a}_{x:\overline{h}|} = E \cdot {}_{h|}\overline{a}_x.$$

Example: Suppose that a person aged x enters into the following life insurance contract: contract term is n, premium payment period is h and the contract consists of a pure endowment insurance $(\mathbf{V}/S/n)$ complemented with a *return-of-premium policy* such that in case the insured dies during the contract term, the paid premiums (without any interest) are paid to the beneficiary. In this case the continuous level premium can be solved from

$$\overline{P} \cdot \overline{a}_{x:\overline{h}|} = S \cdot A_{x:\overline{n}|}(\mathbf{V}) + \overline{P} \cdot (I\overline{A})_{x:\overline{h}|}(\mathbf{K}) + h \cdot \overline{P} \cdot {}_{h|}A_{x:\overline{n-h}|}(\mathbf{K}),$$

which yields

$$\overline{P} = \frac{S \cdot A_{x:\overline{n}|}(\mathbf{V})}{\overline{a}_{x:\overline{h}|} - (I\overline{A})_{x:\overline{h}|}(\mathbf{K}) - h \cdot {}_{h|}A_{x:\overline{n-h}|}(\mathbf{K})},$$

where $(I\overline{A})_{x:\overline{h}|}(\mathbf{K})$ is the expected present value of a linearly growing term life contract in which the sum insured is t if the insured dies at age $x+t$ with $t \le h$. This can be computed as

$$\begin{aligned}
(I\overline{A})_{x:\overline{h}|}(\mathbf{K}) &:= \int_0^h t \cdot {}_tp_x \cdot e^{-\delta t} \cdot \mu_{x+t} dt \\
&= \tfrac{1}{D_x} \int_0^h t \cdot D_{x+t} \cdot \mu_{x+t} dt \\
&= \tfrac{1}{D_x} \cdot \left(-\Big/_0^h t \cdot \overline{M}_{x+t} + \int_0^h \overline{M}_{x+t} dt \right) \\
&= \tfrac{1}{D_x} \cdot \left(-h \cdot \overline{M}_{x+h} + \int_0^h D_{x+t} dt - \delta \cdot \int_0^h \overline{N}_{x+t} dt \right) \\
&= \overline{a}_{x:\overline{h}|} - \left(h \cdot \tfrac{\overline{M}_{x+h}}{D_x} + \tfrac{\delta}{D_x} \int_0^h \overline{N}_{x+t} dt \right).
\end{aligned}$$

2.3.5. *Multiple Life Insurance.* An insurance contract can be dependent on the survival or death of more than one person, i.e. there may be several insured persons. This is called *multiple life insurance* or *joint life insurance* . Typical example is widow's pension, where the widow is paid after the death of the spouse a pension for as long as the widow is alive.

Consider an insurance contract with two insured persons aged x and y. Their remaining lifetimes are $\tilde{T}_x$ and $\tilde{T}_y$. The random variable

$$\tilde{T}_{xy} := \min(\tilde{T}_x, \tilde{T}_y) \tag{142}$$

is the remaining lifetime until first death.

For simplicity, assume that lifetimes $\tilde{T}_x$ and $\tilde{T}_y$ are mutually independent, even though this may not hold in practice (e.g. lifetimes of a married couple may not be independent). By virtue of independence,

$$\begin{aligned} {}_tp_{xy} &:= \mathbb{P}(\tilde{T}_{xy} > t) = \mathbb{P}\left(\{\tilde{T}_x > t\} \cap \{\tilde{T}_y > t\}\right) \\ &= {}_tp_x \cdot {}_tp_y = \exp\left(-\int_0^t (\mu_{x+s} + \mu_{y+s}) ds\right), \end{aligned}$$

and hence the *joint mortality* corresponding to remaining lifetime $\tilde{T}_{xy}$ is

$$\mu_{xy+t} = \mu_{x+t} + \mu_{y+t}. \tag{143}$$

This equation resembles the equations derived when dealing with competing causes of death: in a sense, the insured persons' collective (*joint life status*) dies as its first member dies. Mathematically, joint life insurance contracts are similar to mortality models for several causes of death.

Example: In a unit *widow's pension* (notation: $(\mathbf{F}/1)$) the beneficiary Y is paid a continuous unit pension (i.e. annual amount is $E = 1$) for as long as Y lives after the death of the insured person X. Let the ages of the insured and the beneficiary be x and y at initiation of the contract. To compute the expected present value of this widow's pension, we use the indicators of being alive after time t has passed from initiation of the contract, $\tilde{J}_x(t)$ and $\tilde{J}_y(t)$. Because the pension is paid when X is dead and Y is alive, the net single premium is (by virtue of assumed independence of lifetimes)

$$\begin{aligned} \mathbb{E}\left[\int_0^\infty (1 - \tilde{J}_x(t)) \cdot \tilde{J}_y(t) \cdot e^{-\delta t} dt\right] &= \int_0^\infty \mathbb{E}\left[\tilde{J}_y(t)\right] \cdot e^{-\delta t} dt - \int_0^\infty \mathbb{E}\left[\tilde{J}_y(t)\right] \mathbb{E}\left[\tilde{J}_x(t)\right] \cdot e^{-\delta t} dt \\ &= \int_0^\infty {}_tp_y \cdot e^{-\delta t} dt - \int_0^\infty {}_tp_y \cdot {}_tp_x \cdot e^{-\delta t} dt = \overline{a}_y - \overline{a}_{xy}, \end{aligned}$$

where $\overline{a}_{xy} := \int_0^\infty {}_tp_{xy} \cdot e^{-\delta t} dt$ is the expected present value of a continuous unit life annuity corresponding to joint mortality (143), in which pension is paid until the first death.

Of course, the net single premium could be deduced directly as follows: subtract from the expected present value of a unit whole life annuity paid to the (future) widow beginning from initiation of the contract the expected present value of a unit whole life annuity paid to the joint life status from initiation of the contract – the joint life status is "alive"for as long as both insured persons are alive, and this is also the period from which the (future) widow does not yet receive pension.

Example: In a *orphan's pension* (notation: $(\mathbf{D}/E/n)$) the situation is similar to widow's pension otherwise, but beneficiary Y is replaced by beneficiary Z aged z, to whom pension is paid after the death of the insured X at most until age $w = z+n$. The net single premium is

$$\overline{a}_{z:\overline{n}|} - \overline{a}_{xz:\overline{n}|},$$

where $\overline{a}_{xz:\overline{n}|} := \int_0^n {}_tp_{xz} \cdot e^{-\delta t} dt$. Observe that $(\mathbf{F}/E) = (\mathbf{D}/E/\infty)$.

Calculations with multiple life statuses of more than two lives are in principle similar to the two lives case.

Example: Three persons X, Y, and Z, aged x, y and z, enter the following insurance contract: after the death of X, Y is paid a continuous annual pension of 1 and Z is paid a continuous annual pension of 0.2. If also Y dies, the continuous annual pension paid to Z rises to 0.6. Pension is paid to Y for whole remaining lifetime and to Z at most until age $w = z + n$. Net single premium is

$$\mathbb{E}\big[\int_0^n \{1.2 \cdot (1 - \tilde{J}_x(t)) \cdot \tilde{J}_y(t) \cdot \tilde{J}_z(t) \cdot e^{-\delta t} +$$

$$+1.0 \cdot (1 - \tilde{J}_x(t)) \cdot \tilde{J}_y(t) \cdot (1 - \tilde{J}_z(t)) \cdot e^{-\delta t} +$$

$$+0.6 \cdot (1 - \tilde{J}_x(t)) \cdot (1 - \tilde{J}_y(t)) \cdot \tilde{J}_z(t) \cdot e^{-\delta t}\} dt +$$

$$+ \int_n^\infty (1 - \tilde{J}_x(t)) \cdot \tilde{J}_y(t) \cdot e^{-\delta t} dt\big] =$$

$$= 1.2 \cdot \int_0^n {}_tq_x \cdot {}_tp_y \cdot {}_tp_z \cdot e^{-\delta t} dt + 1.0 \cdot \int_0^n {}_tq_x \cdot {}_tp_y \cdot {}_tq_z \cdot e^{-\delta t} dt +$$

$$+0.6 \cdot \int_0^n {}_tq_x \cdot {}_tq_y \cdot {}_tp_z \cdot e^{-\delta t} dt + 1.0 \cdot \int_n^\infty {}_tq_x \cdot {}_tp_y \cdot e^{-\delta t} dt.$$

Inserting now ${}_tq_h = 1 - {}_tp_h$ into the above formula yields

$$1.2 \cdot \left(\overline{a}_{yz:\overline{n}|} - \overline{a}_{xyz:\overline{n}|}\right) + 1.0 \cdot \left(\overline{a}_{y:\overline{n}|} - \overline{a}_{xy:\overline{n}|} - \overline{a}_{yz:\overline{n}|} + \overline{a}_{xyz:\overline{n}|}\right) +$$

$$+0.6 \cdot \left(\overline{a}_{z:\overline{n}|} - \overline{a}_{xz:\overline{n}|} - \overline{a}_{yz:\overline{n}|} + \overline{a}_{xyz:\overline{n}|}\right) + (\overline{a}_y - \overline{a}_{xy}) - \left(\overline{a}_{y:\overline{n}|} - \overline{a}_{xy:\overline{n}|}\right).$$

Combining terms gives

$$(\overline{a}_y - \overline{a}_{xy}) + 0.6 \cdot \left(\overline{a}_{z:\overline{n}|} - \overline{a}_{xz:\overline{n}|}\right) - 0.4 \cdot \left(\overline{a}_{yz:\overline{n}|} - \overline{a}_{xyz:\overline{n}|}\right).$$

First term corresponds to widow's pension to Y; the second and third terms correspond to such an orphan's pension where we subtract from the orphan's pension ($\mathbf{D}/0.6/n$) an orphan's pension paid to Z after X if Y is alive. If only X is dead but Y is alive, then the pension paid to Z is (in annual amounts) $0.6 - 0.4 = 0.2$.

Consider an insurance contract involving k insured persons $(X_1, \dots, X_n)$, aged $\underline{x} := (x_1, \dots, x_k)$ at initiation of the contract ($\underline{x}$ is multiple joint life status). Denote

$$\tilde{T}_{\underline{x}} := \min_i(\tilde{T}_{x_i}),$$

$$ {}_tp_{\underline{x}} := \Pi_{i=1}^k {}_tp_{x_i} = \mathbb{P}(\tilde{T}_{\underline{x}} > t),$$

$$ {}_{\underline{x}}p_t := \Pi_{i=1}^k {}_{x_i}p_t,$$

$$\mu_{\underline{x}+t\cdot\underline{1}} := \sum_{i=1}^k \mu_{x_i+t},$$

$$D_{\underline{x}} := {}_{\underline{x}}p_0 \cdot e^{\delta\cdot\overline{x}}, \text{ where } \overline{x} = \frac{1}{k}\sum_{i=1}^k x_i,$$

$$\overline{N}_{\underline{x}} := \int_0^\infty D_{\underline{x}+t\cdot\underline{1}} dt,$$

$$\overline{M}_{\underline{x}} := \int_0^\infty D_{\underline{x}+t\cdot\underline{1}} \cdot \mu_{\underline{x}+t\cdot\underline{1}} dt,$$

where $\underline{1} = (1, \dots, 1)$. Using these quantities, the formulas for single life annuities and insurance contracts generalize for multiple life annuities and insurance contracts. In particular, we have the following results (where annuities are paid until the first death).

i: $\overline{a}_{\underline{x}} = \frac{\overline{N}_{\underline{x}}}{D_{\underline{x}}}$;

ii: $\overline{a}_{\underline{x}:\overline{n}|} = \frac{\overline{N}_{\underline{x}} - \overline{N}_{\underline{x}+n\cdot\underline{1}}}{D_{\underline{x}}}$;

iii: $A_{\underline{x}:\overline{n}|}(\mathbf{K}) = \frac{\overline{M}_{\underline{x}} - \overline{M}_{\underline{x}+n\cdot\underline{1}}}{D_{\underline{x}}} = 1 - \delta \cdot \overline{a}_{\underline{x}:\overline{n}|}$.

Proof of (i)-(iii) is left as an exercise.

2.4. **Technical Provisions.** We have previously considered *provisions*. Provisions for insurance liabilities are called *technical provisions*[49] and form a very significant part of an insurer's balance sheet. By differentiation of the provision we can obtain a simple differential equation which facilitates deriving useful formulas for expected present values of insurance liabilities, different provisions and premiums.

Methods considered here are based on claim intensities, which are expected values. Hence these methods do not allow us to consider the probability distributions of derived quantities. In this section we will not consider the corresponding stochastic differential equation, which would allow such considerations.

Technical provisions can be divided into two components: one related to claim events that have already occurred and one related to claim events that will occur in the future. The first component is called *outstanding claim provision*[50]. This provision contains the expected present values of all claim payments arising from claims that have occurred in the past which have not yet been paid. In this course we will concentrate on the second

[49] technical provision = ansvarskuld = vastuuvelka

[50] outstanding claim provision = ersättningsansvar = korvausvastuu

component, the *mathematical provision*[51], which contains the expected present values of claim payments from claim events that will occur in the future.

Mathematical provision can be calculated *retrospectively*, as the expected present value of cumulative collected premiums less the expected present value of claim intensities. Retrospective provision (liability) is hence the result of premiums and claim payments not being simultaneous. *Prospective provision*, in contrast, is the expected present value of future claim payments less expected present value of future insurance premiums. Prospective provision is hence the amount by which the expected present value of future claim costs exceeds the expected present value of future premiums.

It turns out that both calculation methods lead to same result, if the calculation bases for premiums and provisions (claim intensities, interest rate assumptions and loadings) are identical.

To shorten the notations, we adopt the following conventions:

$$A_{x:w}(\cdot) := A_{x:\overline{n}|}(\cdot),$$

where $w = x + n$ is the termination age of the contract, and

$$A_{x:w}(\cdot) = a_{x:w} = 0, \text{ if } x > w.$$

Correspondingly, we agree (in contrast to usual conventions in calculus) that the value of an integral is zero if the lower bound of integration is larger that the upper bound of integration.

2.4.1. *Prospective Provision.* Prospective provision is the difference of the expected present values of future claim costs and of future insurance premiums.

Expected present values of future claim costs for different insurance contracts were derived in previous sections. These net single premiums $S \cdot A_{x:w}(\cdot)$ are the expected present values of future claim costs at the beginning of the contract when insured is aged x, and t years later the expected present value of future claim payments is $S \cdot A_{x+t:w}(\cdot)$. The expected present value of future premiums $\overline{P} \cdot \overline{a}_{x:y}$ has also been derived, and t years later this value is $\overline{P} \cdot \overline{a}_{x+t:y}$ (here $y = x + m$, where m is the premium payment term).

Definition 2.4.1: The *prospective provision* of an insurance contract during contract term is

$$V_t := S \cdot A_{x+t:w}(\cdot) - \overline{P} \cdot \overline{a}_{x+t:y}. \tag{144}$$

Observe that for $x + t > y$ the second term is zero.

Example: Prospective provisions for term life insurance ($\mathbf{K}/S/n$) and pure endowment insurance ($\mathbf{V}/S/n$) can be expressed in terms of commutation functions as follows:

$$V_t(\mathbf{K}) = S \cdot \frac{\overline{M}_{x+t} - \overline{M}_w}{D_{x+t}} - \overline{P} \cdot \frac{\overline{N}_{x+t} - \overline{N}_y}{D_{x+t}}$$

[51]mathematical provision = premieansvar = vakuutusmaksuvastuu

and

$$V_t(\mathbf{V}) = S \cdot \frac{D_w}{D_{x+t}} - \overline{P} \cdot \frac{\overline{N}_{x+t} - \overline{N}_y}{D_{x+t}}.$$

Example: Prospective provision for a pension insurance where the insurance is taken at age x and annual pension paid from age $w = x + n$ until age v equals E, can be expressed in terms of commutation functions as follows:

$$V_t = E \cdot \frac{D_w}{D_{x+t}} \cdot \overline{a}_{w:v} - \overline{P} \cdot \frac{\overline{N}_{x+t} - \overline{N}_y}{D_{x+t}}.$$

2.4.2. *Retrospective Provision.* Retrospective provision is the difference of the expected present values of paid premiums and of paid claim payments.

Expected present value of premiums paid until time t at the beginning of contract is $\overline{P} \cdot \overline{a}_{x:\overline{t}|}$, and accumulating this to time t we obtain the expected present value of premiums paid until time t at time t as

$$\overline{P} \cdot \overline{a}_{x:\overline{t}|} \cdot \frac{D_x}{D_{x+t}}.$$

The expected present value of past claim payments made until time t is $S \cdot A_{x:\overline{t}|}(\cdot)$, and hence their expected present value at time t is

$$S \cdot A_{x:\overline{t}|}(\cdot) \cdot \frac{D_x}{D_{x+t}}.$$

Definition 2.4.2: The *retrospective provision* of an insurance contract during premium payment term $(t \leq y)$ is

$$V_t^{retro} := \overline{P} \cdot \overline{a}_{x:\overline{t}|} \cdot \frac{D_x}{D_{x+t}} - S \cdot A_{x:\overline{t}|}(\cdot) \cdot \frac{D_x}{D_{x+t}}; \tag{145}$$

after the premium payment term $(t > y)$ the retrospective provision is

$$V_t^{retro} := \overline{P} \cdot \overline{a}_{x:\overline{ty}|} \cdot \frac{D_x}{D_{x+t}} - S \cdot A_{x:\overline{t}|}(\cdot) \cdot \frac{D_x}{D_{x+t}}, \tag{146}$$

i.e. only premium payment period is taken into account in the term related to paid premiums.

Example: Retrospective provision for term life insurance $(\mathbf{K}/S/n)$ can be expressed in terms of commutation functions as follows:

$$V_t^{retro}(\mathbf{K}) = \overline{P} \cdot \frac{\overline{N}_x - \overline{N}_{x+t}}{D_{x+t}} - S \cdot \frac{\overline{M}_x - \overline{M}_{x+t}}{D_{x+t}}.$$

To establish the equality of retrospective and prospective provisions, we need the following proposition:

Proposition: The following relations hold for life annuities and expected present values of life insurance contracts:

i: $\overline{a}_{x:y} = \overline{a}_{x:\overline{t}|} + \frac{D_{x+t}}{D_x} \cdot \overline{a}_{x+t:y}$;

ii: $A_{x:w}(\mathbf{K}) = A_{x:\overline{t}|}(\mathbf{K}) + \frac{D_{x+t}}{D_x} \cdot A_{x+t:w}(\mathbf{K})$;

iii: $A_{x:w}(\mathbf{V}) = \frac{D_{x+t}}{D_x} \cdot A_{x+t:w}(\mathbf{V}) = A_{x:x+t}(\mathbf{V}) \cdot A_{x+t:w}(\mathbf{V})$;

With the help of this proposition, it is possible to prove the following.

Theorem: Prospective provision equals retrospective provision for each insurance contract type considered (i.e. term life $\mathbf{K}$, pure endowment $\mathbf{V}$ and endowment life $\mathbf{Y}$), if the applied calculation base is the same.

Proofs of both this theorem and the preceding auxiliary proposition are left as exercises.

2.5. Thiele's Differential Equation.

2.5.1. *Basic Form of Thiele's Equation.* We shall next derive Thiele's differential equation with the formula for prospective provision (144) as our starting point. Multiplying equation (144) with D_{x+t} and differentiating with respect to t, we obtain Thiele's equation

$$\frac{d}{dt}\left(V_t \cdot D_{x+t}\right) = S \cdot \frac{d}{dt}\left(A_{x+t:w}(\mathbf{K}) \cdot D_{x+t}\right) - \overline{P} \cdot \frac{d}{dt}\left(\overline{a}_{x+t:y} \cdot D_{x+t}\right). \tag{147}$$

Using previously proved representation results, we can compute more explicit expressions for the derivatives in equation (147):

$$\tfrac{d}{dt}\left(V_t \cdot D_{x+t}\right) = V_t' \cdot D_{x+t} - (\delta + \mu_{x+t}) \cdot D_{x+t} \cdot V_t;$$

$$\tfrac{d}{dt}\left(A_{x+t:w}(\mathbf{K}) \cdot D_{x+t}\right) = \tfrac{d}{dt}\left(\overline{M}_{x+t} - \overline{M}_w\right) = \tfrac{d}{dt}\left(\overline{M}_{x+t}\right) = -\mu_{x+t} \cdot D_{x+t};$$

$$\tfrac{d}{dt}\left(\overline{a}_{x+t:y} \cdot D_{x+t}\right) = \tfrac{d}{dt}\left(\overline{N}_{x+t} - \overline{N}_y\right) = \tfrac{d}{dt}\left(\overline{N}_{x+t}\right) = -D_{x+t}$$

Inserting these expressions into equation (147) and dividing by D_{x+t} we obtain the basic form of *Thiele's differential equation*

$$V_t' - (\delta + \mu_{x+t}) \cdot V_t = -S \cdot \mu_{x+t} + \overline{P}. \tag{148}$$

Rearranging the terms yields

$$V_t' = \delta \cdot V_t + \mu_{x+t} \cdot V_t + \overline{P} - \mu_{x+t} \cdot S, \tag{149}$$

which gives us a decomposition of the growth V_t' of the provision into different components: the insurer earns interest (investment income) on the assets backing the provision, and from this interest the provision is credited with amount $\delta \cdot V_t$; the insurer receives premiums $\overline{P}$, which are added to the provision; upon death of the insured, the insurer pays the sum insured S, and the corresponding intensity $\mu_{x+t} \cdot S$ is deducted from the provision; in addition, upon the death of the insured the insurance terminates and the insurer's liabilities decrease by V_t, and hence the corresponding intensity $\mu_{x+t} \cdot V_t$ is credited to the provision (so-called *mortality bonus*).

While the decomposition (149) was derived for term life insurance, it can be applied to pure endowment and endowment life contracts by choosing the boundary conditions of the differential equation in a way consistent with the terms of the insurance contract type.

If a single premium P_0 is paid at beginning of the contract term, then the initial condition is set to $V_0 = P_0$; otherwise we set $V_0 = 0$. Similarly, if it is agreed that a sum S will be paid at the termination of the insurance contract provided that the insured is then alive, we set the terminal condition to $V_n = S$; otherwise we set $V_n = 0$.

For a term life insurance contract with continuous payments ($m = n$), Thiele's equation with boundary conditions is

$$\begin{cases} V_t' = \delta \cdot V_t + \mu_{x+t} \cdot V_t + \overline{P} - \mu_{x+t} \cdot S \\ V_0 = 0,\ V_n = 0, \end{cases} \tag{150}$$

and for a term life insurance contract with a single premium P_0 paid at the beginning

$$\begin{cases} V_t' = \delta \cdot V_t + \mu_{x+t} \cdot V_t - \mu_{x+t} \cdot S \\ V_0 = P_0,\ V_n = 0. \end{cases} \tag{151}$$

For a pure endowment insurance with continuous payments ($m = n$), Thiele's equation with boundary conditions is

$$\begin{cases} V_t' = \delta \cdot V_t + \mu_{x+t} \cdot V_t + \overline{P} \\ V_0 = 0,\ V_n = S, \end{cases} \tag{152}$$

and for a pure endowment insurance contract with a single premium P_0 paid at the beginning

$$\begin{cases} V_t' = \delta \cdot V_t + \mu_{x+t} \cdot V_t \\ V_0 = P_0,\ V_n = S. \end{cases} \tag{153}$$

For a endowment life insurance with continuous payments ($m = n$), Thiele's equation with boundary conditions is

$$\begin{cases} V_t' = \delta \cdot V_t + \mu_{x+t} \cdot V_t + \overline{P} - \mu_{x+t} \cdot S \\ V_0 = 0,\ V_n = S, \end{cases} \tag{154}$$

and for a endowment life insurance contract with a single premium P_0 paid at the beginning

$$\begin{cases} V_t' = \delta \cdot V_t + \mu_{x+t} \cdot V_t - \mu_{x+t} \cdot S \\ V_0 = P_0,\ V_n = S. \end{cases} \tag{155}$$

2.5.2. *Generalizations of Thiele's Equation.* The conditions of the insurance contract determine the terms in Thiele's equation. Following generalizations increase the applicability of equation (149) to insurance contracts with different characteristics.

(1) In a term life insurance contract, the sum insured payable upon death may be specified to change in time: hence we introduce the sum insured S_t which depends (deterministically) on time. Then Thiele's equation takes the form

$$\begin{cases} V_t' = \delta \cdot V_t + \mu_{x+t} \cdot V_t + \overline{P} - \mu_{x+t} \cdot S_t \\ V_0 = P_0,\ V_n = 0. \end{cases} \tag{156}$$

(2) In a pure endowment insurance, the sum insured is paid at the termination of the contract provided that the insured is alive then. It is possible that additionally a sum insured is paid at some earlier age provided that the insured is then alive; it is also possible that the sum insured of the pure endowment insurance part of an insurance contract is agreed to be paid before the termination age of some other parts of the contract. To accommodate such features, we add to the right side of Thiele's equation a new term $-E_t$ containing all sums insured conditional on survival. Now the terminal condition must be set to $V_n = 0$, since the sum insured is contained in the new term, and Thiele's equation takes the form

$$\begin{cases} V_t' = \delta \cdot V_t + \mu_{x+t} \cdot V_t + \overline{P} - \mu_{x+t} \cdot S_t - E_t \\ V_0 = P_0,\, V_n = 0. \end{cases} \tag{157}$$

(3) To accommodate a premium changing in time, we introduce the deterministic time-dependent premium P_t, in which case Thiele's equation is

$$\begin{cases} V_t' = \delta \cdot V_t + \mu_{x+t} \cdot V_t + P_t - \mu_{x+t} \cdot S_t - E_t \\ V_0 = P_0,\, V_n = 0. \end{cases} \tag{158}$$

(4) If the sum insured is payable upon death, then claim intensity equals $\mu_{x+t} \cdot S$. Sum insured may be payable upon some other event, such as illness or disability. If the intensity associated with this insured event is ν_t, then the claim intensity $\mu_{x+t} \cdot S$ in previous equations must be replaced with $\nu_t \cdot S$. This gives Thiele's equation

$$\begin{cases} V_t' = \delta \cdot V_t + \mu_{x+t} \cdot V_t + P_t - \nu_t \cdot S_t - E_t \\ V_0 = P_0,\, V_n = 0. \end{cases} \tag{159}$$

(5) An insurance contract may stipulate that sum insured is payable upon many different events, e.g. both upon death and disability. The sums insured may also be different in each event. Denote the vector of relevant claim intensities $\underline{\nu}_t := (\nu_{1,t}, \nu_{2,t}, \ldots, \nu_{k,t})$ and the vector of sums insured $\underline{S}_t := (S_{1,t}, S_{2,t}, \ldots, S_{k,t})$, where k is the number of different risks covered and $\nu_{j,t}$ is the claim intensity associated with risk j at time t. Total claim intensity is then the sum $\sum_{i=1}^{k} \nu_{i,t} \cdot S_{i,t} = \underline{\nu}_t \cdot \underline{S}_t$ and Thiele's equation is

$$\begin{cases} V_t' = \delta \cdot V_t + \mu_{x+t} \cdot V_t + P_t - \underline{\nu}_t \cdot \underline{S}_t - E_t \\ V_0 = P_0,\, V_n = 0. \end{cases} \tag{160}$$

(6) Mortality bonus $\mu_{x+t} \cdot V_t$ can be generalized to other reasons for termination of the contract besides death. The mortality bonus is added to the provision and is meant to be financed by such provisions which by contract terms upon the death of the insured remain on the insurer's balance sheet. However, similar agreement can be made concerning also other reasons of contract termination. If the intensities of different such reasons (events) are gathered into a vector $\underline{\tau}_t = (\tau_{1,t}, \tau_{2,t}, \ldots, \tau_{k,t})$,

the bonus term is $\sum_{i=1}^{k} \tau_{i,t} \cdot V_t = \underline{\tau}_t \cdot V_t$ and Thiele's equation is

$$\begin{cases} V_t' = \delta \cdot V_t + \underline{\tau}_t \cdot V_t + P_t - \underline{\nu}_t \cdot \underline{S}_t - E_t \\ V_0 = P_0,\, V_n = 0. \end{cases} \tag{161}$$

(7) Force of interest could also be allowed to change in time, in which case Thiele's equation is

$$\begin{cases} V_t' = \delta_t \cdot V_t + \underline{\tau}_t \cdot V_t + P_t - \underline{\nu}_t \cdot \underline{S}_t - E_t \\ V_0 = P_0,\, V_n = 0. \end{cases} \tag{162}$$

For the derived generalized Thiele's differential equation, we also need to generalize the commutation functions. Observing that terms related to interest $\delta_t \cdot V_t$ and termination bonus $\underline{\tau}_t \cdot V_t$ are directly related to the size of the provision, we define the generalized commutation function

$$D_t := \exp\left(- \int_0^t (\delta_s + \underline{\tau}_s) ds \right). \tag{163}$$

Generalizations of other commutation functions are derived from D_x:

$$\overline{N}_x := \int_x^\infty D_t dt \text{ and } \overline{M}_x^{(i)} := \int_x^\infty D_t \cdot \nu_{i,t} dt, \tag{164}$$

where $\overline{M}_x^{(i)}$ is defined separately for each claim intensity $\nu_{i,t}$. Based on these commutation functions, the net single premium for a unit pure endowment insurance for a person aged x terminating at age w is

$$A_{x:w}(\mathbf{V}) := \frac{D_w}{D_x}.$$

Similarly, the expected present value of premiums is

$$\overline{a}_{x:y} := \frac{\overline{N}_x - \overline{N}_y}{D_x}$$

and the net single premium for each insurance corresponding to a claim intensity is

$$A_{x:w}(i) := \frac{\overline{M}_x^{(i)} - \overline{M}_w^{(i)}}{D_x}.$$

2.5.3. *Equivalence Equation.* Next we will derive the equivalence equation from which formulas for premiums and provisions can be derived, by integrating the one-dimensional generalized Thiele's differential equation (162). Multiplying equation (162) by D_t and integrating over $(a,b) \subseteq (0,n)$, we obtain

$$\int_a^b \left(V_t' \cdot D_t - (\delta + \tau_t) \cdot D_t \cdot V_t \right) dt = \int_a^b \left(P_t - E_t - \nu_t \cdot S_t \right) D_t dt. \tag{165}$$

The left side of above equation can be written as

$$\int_a^b \left(V_t' \cdot D_t - (\delta + \tau_t) \cdot D_t \cdot V_t \right) dt = \int_a^b \frac{d}{dt}(D_t \cdot V_t) dt = D_b \cdot V_b - D_a \cdot V_a,$$

and hence we can write equation (165) as

$$D_b \cdot V_b - D_a \cdot V_a = \int_a^b P_t \cdot D_t dt - \int_a^b E_t \cdot D_t dt - \int_a^b \nu_t \cdot S_t \cdot D_t dt. \tag{166}$$

This means that the change in the expected present value of the provision between times a and b equals the expected present value of paid premiums during that time interval less the expected present value of claim intensities during the same time interval.

We also need to set the boundary conditions. Initial provision V_0 can be chosen to be either zero or equal to the possible single premium paid at beginning. In the first case, the possible initial single premium is included in the premium term $\int_0^n P_t \cdot D_t dt$, while in the latter case the initial premium is not included in the term. Similarly, the final provision can be chosen to equal zero or to equal the possible sum insured payable at termination of the contract. In the first case, the possible sum insured payable is included in the term $\int_0^n E_t \cdot D_t dt$, while in the latter case this sum insured is not included in the term. In this presentation we choose $V_0 = V_n = 0$, and hence integrating equation (166) over $(0, n)$ yields the *equivalence equation*

$$\int_0^n P_t \cdot D_{x+t} dt = \int_0^n E_t \cdot D_{x+t} dt + \int_0^n \nu_x + t \cdot S_t \cdot D_{x+t} dt. \tag{167}$$

Equivalence equation states that the expected present value of premiums equals the expected present value of claim intensities, and hence the cash flows of premiums and of claims are equivalent.

In practice, premium payments and payments of sums insured are discrete, in which case the integrals must be replaced with summations.

2.5.4. *Premiums.* Formulas for premiums can be derived by dividing the equivalence equation (167) by D_x, which leads to

$$\int_0^n P_t \cdot \frac{D_{x+t}}{D_x} dt = \int_0^n E_t \cdot \frac{D_{x+t}}{D_x} dt + \int_0^n \nu_x + t \cdot S_t \cdot \frac{D_{x+t}}{D_x} dt. \tag{168}$$

If premium is a constant $\overline{P}$, the payment term ends at age y, sum insured is S and a single sum insured S_w is paid upon termination, then equation (168) simplifies to

$$\overline{P} \cdot \overline{a}_{x:y} = S_w \cdot A_{x:w}(\mathbf{V}) + S \cdot A_{x:w}(\cdot) \tag{169}$$

or, in terms of commutation functions,

$$\overline{P} \cdot \frac{\overline{N}_x - \overline{N}_{x+m}}{D_x} = S_w \cdot \frac{D_w}{D_x} + S \cdot \frac{\overline{M}_x - \overline{M}_{x+n}}{D_x}, \tag{170}$$

from which we obtain expressions for premiums as

$$\overline{P} = \frac{S_w \cdot A_{x:w}(\mathbf{V}) + S \cdot A_{x:w}(\cdot)}{\overline{a}_{x:y}} \tag{171}$$

and

$$\overline{P} = \frac{S_w \cdot D_w + S \cdot (\overline{M}_x - \overline{M}_{x+n})}{\overline{N}_x - \overline{N}_{x+m}}. \tag{172}$$

2.5.5. *Technical Provision.* At initiation of the insurance contract, the insurer's liability for claims begins and the insurer begins to receive the agreed premiums. Future claim payments are liabilities for the insurer and for accounting purposes are classified as insurer's debt to policyholders. This liability is called the technical provision, which the insurer must calculate annually. Previously derived formulas for provisions of term life, pure endowment and pension insurance contracts remain valid after generalizations made to Thiele's equation in section 2.5.2.

By integrating in equation (166) over (t, n) we obtain the equation for calculating prospective provision

$$D_n \cdot V_n - D_t \cdot V_t = \int_t^n P_t \cdot D_t dt - \int_t^n E_t \cdot D_t dt - \int_t^n \nu_t \cdot S_t \cdot D_t dt. \tag{173}$$

Since the final provision $V_n = 0$, this can be solved for prospective provision V_t:

$$V_t = \int_t^n E_u \cdot \frac{D_u}{D_t} dt - \int_t^n \nu_u \cdot S_u \cdot \frac{D_u}{D_t} dt - \int_t^n P_u \cdot \frac{D_u}{D_t} dt. \tag{174}$$

This could also be obtained directly from the one-dimensional Thiele's differential equation (162) with final condition $V_n = 0$, using general solution formulas for non-homogeneous first order linear differential equations.

If premium is a constant $\overline{P}$, the payment term ends at age y, sum insured is S and a single sum insured S_w is paid upon termination, then the above equation is often written (when $x + t < y$) as

$$V_t = S_w \cdot A_{x+t:w}(\mathbf{V}) + S \cdot A_{x+t:w}(\cdot) - \overline{P} \cdot \overline{a}_{x+t:y}, \tag{175}$$

or, in terms of commutation functions, as

$$V_t = S_w \cdot \frac{D_w}{D_{x+t}} + S \cdot \frac{\overline{M}_{x+t} - \overline{M}_w}{D_{x+t}} - \overline{P} \cdot \frac{\overline{M}_{x+t} - \overline{M}_w}{D_{x+t}}. \tag{176}$$

For $x + t \geq y$, the last term (premiums) vanishes.

In similar fashion, by integrating in equation (166) over $(0, t)$ we obtain the retrospective provision

$$V_t^{retro} = \int_0^t P_u \cdot \frac{D_u}{D_t} du - \int_0^t E_u \cdot \frac{D_u}{D_t} du - \int_0^t S_u \cdot \nu_u \cdot \frac{D_u}{D_t} du, \tag{177}$$

and it has representations

$$V_t^{retro} = \overline{P} \cdot \overline{a}_{x:x+t} \cdot \frac{D_x}{D_{x+t}} - S \cdot A_{x:x+t} \cdot \frac{D_x}{D_{x+t}} = \overline{P} \cdot \frac{\overline{N}_x - \overline{N}_{x+t}}{D_{x+t}} - S \cdot \frac{\overline{M}_x - \overline{M}_{x+t}}{D_{x+t}}.$$

If the calculation bases for premiums are the same, prospective and retrospective provisions are equal also in the generalized case. Retrospective provision (177) could also be obtained directly from the one-dimensional Thiele's differential equation (162) with initial condition $V_0 = 0$, using general solution formulas for non-homogeneous first order linear differential equations.

Example: For a term life insurance with a constant premium, Thiele's differential equation is

$$V_t' = \delta \cdot V_t + \mu_{x+t} \cdot V_t + \overline{P} - \mu_{x+t} \cdot S,$$

with boundary conditions $V_n = V_0 = 0$. Equivalence equation is

$$\int_0^m \overline{P} \cdot D_{x+t} dt = \int_0^n S \cdot \mu_{x+t} \cdot D_{x+t} dt,$$

where $D_x = \exp\left(-\int_0^x (\delta + \mu_s) ds\right)$. If payment term ends at age y, dividing by D_x we obtain from above equation the form

$$\overline{P} \cdot \overline{a}_{x:y} = \int_0^n S \cdot \mu_{x+t} \cdot \frac{D_{x+t}}{D_x} dt.$$

Prospective provision at time t is during premium payment term

$$V_t = S \cdot A_{x+t:w}(\mathbf{K}) - \overline{P} \cdot \overline{a}_{x+t:y}$$

(after premium payment term the last term is zero).

Example: For a pure endowment insurance, Thiele's differential equation is

$$V_t' = \delta \cdot V_t + \mu_{x+t} \cdot V_t + P_t,$$

with boundary conditions $V_0 = 0$, $V_n = S_w$. Equivalence equation is

$$\int_0^m P_t \cdot D_{x+t} dt = S_w \cdot D_w,$$

where $D_x = \exp\left(-\int_0^x (\delta + \mu_s) ds\right)$. If payment term ends at age y and payment is constant $P_t = \overline{P}$, dividing by D_x we obtain from above equation the form

$$\overline{P} \cdot \overline{a}_{x:y} = S_w \cdot A_{x:w}(\mathbf{V}).$$

Prospective provision at time t is during premium payment term

$$V_t = S_w \cdot A_{x+t:w}(\mathbf{V}) - \overline{P} \cdot \overline{a}_{x+t:y}$$

(after premium payment term the last term is zero).

Example: For a pension insurance, Thiele's differential equation is

$$V_t' = \delta \cdot V_t + \mu_{x+t} \cdot V_t + P_t,$$

with boundary conditions $V_0 = 0$, $V_n = E \cdot \overline{a}_{w:w'}$. Equivalence equation is

$$\int_0^m P_t \cdot D_{x+t} dt = E \cdot \overline{a}_{w:w'} \cdot D_w,$$

where $D_x = \exp\left(-\int_0^x (\delta + \mu_s) ds\right)$. If payment term ends at age y and payment is constant $P_t = \overline{P}$, dividing by D_x we obtain from above equation the form

$$\overline{P} \cdot \overline{a}_{x:y} = E \cdot {}_{n|}\overline{a}_{x:\overline{w'-w}|}.$$

Prospective provision at time t is during premium payment term

$$V_t = E \cdot \overline{a}_{w:w'} \cdot \frac{D_w}{D_{x+t}} - \overline{P} \cdot \overline{a}_{x+t:y}$$

(after premium payment term the last term is zero).

Example: Consider an insurance contract where the insurer pays sum insured S if the contract terminates because of disability, death or attaining of the termination age w. If the disability intensity for a person aged $x+t$ is ν_{x+t}, Thiele's differential equation is

$$V_t' = (\delta + \mu_{x+t} + \nu_{x+t}) \cdot V_t - (\mu_{x+t} + \nu_{x+t}) \cdot S + P_t,$$

with boundary conditions $V_0 = 0$, $V_n = S$. Equivalence equation is

$$\int_0^m P_t \cdot D_{x+t} dt = S \cdot D_w + S \cdot \int_0^n (\mu_{x+t} + \nu_{x+t}) \cdot D_{x+t} dt,$$

where $D_x = \exp\left(-\int_0^x (\delta + \mu_s) ds\right)$. If payment term ends at age y and payment is constant $P_t = \overline{P}$, we obtain from above equation the form

$$\overline{P} \cdot \overline{a}_{x:y} = S \cdot A_{x:w}(\mathbf{V}) + S \cdot A_{x:w},$$

where $A_{x:w} := \int_0^n (\nu_{x+t} + \mu_{x+t}) \cdot \frac{D_w}{D_{x+t}} dt$. Prospective provision at time t is during premium payment term

$$V_t = S \cdot A_{x+t:w}(\mathbf{V}) + S \cdot A_{x+t:w} - \overline{P} \cdot \overline{a}_{x+t:y}$$

(after premium payment term the last term is zero).

Example: Consider an insurance contract where the insurer pays sum insured S when first of the two insured persons dies (notation: **KK**). Now claim intensity for joint life status is $\nu_t = \mu_{x+t} + \mu_{y+t}$, and Thiele's differential equation is

$$V_t' = (\delta + \mu_{x+t} + \mu_{y+t}) \cdot V_t - (\mu_{x+t} + \mu_{y+t}) \cdot S + P_t,$$

with boundary conditions $V_0 = 0$, $V_n = 0$. With constant continuous payment $\overline{P}$, equivalence equation is

$$\overline{P} \cdot \int_0^m D_{xy+t} dt = S \cdot \int_0^n (\mu_{x+t} + \mu_{y+t}) \cdot D_{xy+t} dt,$$

where $D_{xy+t} = \exp\left(-\int_0^t (\delta + \mu_{x+s} + \mu_{y+s}) ds\right)$. If payment term is m, we obtain from above equation the form

$$\overline{P} \cdot \overline{a}_{xy:\overline{m}|} = S \cdot A_{xy:\overline{n}|}(\mathbf{KK}),$$

where $A_{xy:\overline{n}|}(\mathbf{KK}) := \int_0^n (\mu_{x+t} + \mu_{y+t}) \cdot \frac{D_{xy+t}}{D_{xy}} dt$. Prospective provision at time t is during premium payment term

$$V_t = S \cdot A_{xy+t:\overline{n-t}|}(\mathbf{KK}) - \overline{P} \cdot \overline{a}_{xy+t:\overline{m-t}|}$$

(after premium payment term the last term is zero).

2.6. **Expense Loadings.** Previously derived premiums cover the insurer's expected claim costs. In addition to these costs, the insurer incurs operating or administrative expenses, such as wages and social security payments of employees. To cover these expenses, an *expense loading* is added to the insurance premium. We denote net premiums (which exclude expense loadings) by P and *gross premiums* which include expense loadings by B.

Loadings can be *proportional*, e.g.

- **i:** proportional to gross premium, $\kappa \cdot B_t$;
- **ii:** proportional to sum insured, $\epsilon \cdot S_t$;
- **iii:** proportional to claim intensity, $\phi \cdot \nu_t \cdot S_t$
- **iv:** proportional to provision, $\gamma \cdot V_t$;

or *absolute*, such as

- **i:** the expenses incurred when setting up the insurance contract;
- **ii:** constant annual expenses from administration of the contract and claim handling.

Equivalence principle is generalized to gross premiums as follows:

$$\text{PV of gross premiums} = \text{PV of claim costs} + \text{PV of expenses} \tag{178}$$

As with claim costs, in this presentation present value of expenses is computed as an expected value.

Thiele's differential equation is now

$$V_t' = (\delta + \tau_t) \cdot V_t + B_t - E_t - \nu_t \cdot S_t - L_t, \tag{179}$$

where L_t is the expense loading at time t, which can contain both proportional and absolute loadings (see list above). As before, integration over $[a, b] \subseteq [0, n]$ yields

$$D_b V_b - D_a V_a = \int_a^b B_t \cdot D_t dt - \int_a^b E_t \cdot D_t dt - \int_a^b S_t \cdot \nu_t \cdot D_t dt - \int_a^b L_t \cdot D_t dt. \tag{180}$$

One interpretation of this equation is that premiums increase technical provisions to the extent that they are not needed to cover claim costs and operating expenses.

Equivalence equation taking into account expense loadings is

$$\int_0^m B_t \cdot D_{x+t} dt = \int_0^n E_t \cdot D_{x+t} dt + \int_0^n S_t \cdot \nu_{x+t} \cdot D_{x+t} dt + \int_0^n L_t \cdot D_{x+t} dt. \tag{181}$$

For level continuous gross premium $\overline{B}$ and constant sums insured S_w and S we have

$$\overline{B} \cdot \overline{a}_{x:y} = S_w \cdot A_{x:w}(\mathbf{V}) + S \cdot A_{x:w}(\cdot) + \int_0^n L_t \cdot \frac{D_{x+t}}{D_x} dt,$$

from which we can solve the gross premium

$$\overline{B} = \frac{S_w \cdot A_{x:w}(\mathbf{V}) + S \cdot A_{x:w}(\cdot) + \int_0^n L_t \cdot \frac{D_{x+t}}{D_x} dt}{\overline{a}_{x:y}}.$$

Prospective provision takes the form

$$V_t = \int_t^n E_u \cdot \frac{D_u}{D_t} du + \int_t^n S_u \cdot \nu_u \cdot \frac{D_u}{D_t} du + \int_t^n L_u \cdot \frac{D_u}{D_t} du - \int_t^n B_u \cdot \frac{D_u}{D_t} du,$$

or (with constant level gross premium and constant sums insured)

$$V_t = S_w \cdot A_{x+t:w}(\mathbf{V}) + S \cdot A_{x+t:w} + \int_t^n L_s \cdot \frac{D_{x+s}}{D_{x+t}} ds - \overline{B} \cdot \overline{a}_{x+t:y},$$

i.e. prospective provision is the expected present value of future claim costs and expenses less the expected present value of future premiums.

Retrospective provision with expense loadings is

$$V_t^{retro} = \int_0^t B_u \cdot \frac{D_u}{D_t} du - \int_0^t E_u \cdot \frac{D_u}{D_t} du - \int_0^t S_u \cdot \nu_u \cdot \frac{D_u}{D_t} du - \int_0^t L_u \cdot \frac{D_u}{D_t} du,$$

i.e. retrospective provision is the expected present value of past premiums less the expected present value of past claim costs and expenses.

Example: Consider an **Y**-insurance, in which the insurer increases the premiums to cover its operating expenses as follows: during premium payment term, amount $\kappa \cdot B_t$ is charged from contract and during contract term amounts $\epsilon \cdot S$ and $\phi \cdot \mu_{x+t} \cdot S$ are charged from contract. Hence the expense loading is

$$L_t = \kappa \cdot B_t + \epsilon \cdot S + \phi \cdot \mu_{x+t} \cdot S.$$

Thiele's differential equation is

$$V_t' - (\delta + \mu_{x+t}) \cdot V_t = B_t - \mu_{x+t} \cdot S - L_t = (1-\kappa) \cdot B_t - (1+\phi) \cdot \mu_{x+t} \cdot S - \epsilon \cdot S$$

with boundary conditions $V_0 = 0$ and $V_n = S$. From this we get for $[a, b] \subset [0, n]$

$$V_b \cdot D_{x+b} - V_a \cdot D_{x+a} = (1-\kappa) \cdot \int_a^b B_t \cdot D_{x+t} dt - S \cdot D_w - S \cdot \int_a^b ((1+\phi)\mu_{x+t} + \epsilon) \cdot D_{x+t} dt.$$

Equivalence equation is

$$(1-\kappa) \cdot \int_0^n B_t \cdot D_{x+t} dt = S \cdot D_w + S \cdot \int_0^n ((1+\phi)\mu_{x+t} + \epsilon) \cdot D_{x+t} dt,$$

and dividing by D_x and assuming $B_t = \overline{B}$, we obtain equivalence

$$(1-\kappa) \cdot \overline{B} \cdot \overline{a}_{x:y} = S \cdot A_{x:w}(\mathbf{V}) + S \cdot ((1+\phi) A_{x:w}(\mathbf{K}) + \epsilon \cdot \overline{a}_{x:w}),$$

gross premium

$$\overline{B} = \frac{S \cdot A_{x:w}(\mathbf{V}) + S \cdot ((1+\phi) A_{x:w}(\mathbf{K}) + \epsilon \cdot \overline{a}_{x:w})}{(1-\kappa) \cdot \overline{a}_{x:y}}$$

and prospective provision

$$V_t = S \cdot (A_{x+t:w}(\mathbf{V}) + (1+\phi) A_{x+t:w}(\mathbf{K}) + \epsilon \cdot \overline{a}_{x+t:w}) - (1-\kappa) \cdot \overline{B} \cdot \overline{a}_{x+t:y}.$$

2.7. Some Special Issues in Life Insurance Contracts.

2.7.1. *Surrender Value and Zillmerization.* Insurance contract may stipulate that the policyholder has the right (option) to terminate the contract before the end of the contract term. In this case the insured sells the contract back to the insurer and the insurer buys the contract and pays its *surrender value*[52] to the policyholder. Surrender value is at most the technical provision of the contract.

Since the insurer has paid as operating expenses the acquisition costs of the contract, policyholder's exercise of the surrender option may cause losses to the insurer depending on how the acquisition costs are amortized from expense loadings. In case an absolute initial loading equal to actual acquisition costs is used, no losses will arise. In practice, however, acquisition costs are often covered instead by increasing other loadings. Suppose that in order to cover acquisition costs I, the loading proportional to gross premiums is increased from κ to κ' in such a way that

$$I = (\kappa' - \kappa) \cdot \overline{B} \cdot \overline{a}_{x:y}.$$

The provision is for loadings $(0, \kappa')$

$$S \cdot ((1 + \phi) \cdot A_{x+t:w} + \epsilon \cdot \overline{a}_{x+t:w}) - (1 - \kappa') \cdot \overline{B} \cdot \overline{a}_{x+t:y}.$$

If loadings (I, κ) had been used, the provision would be

$$S \cdot ((1 + \phi) \cdot A_{x+t:w} + \epsilon \cdot \overline{a}_{x+t:w}) - (1 - \kappa) \cdot \overline{B} \cdot \overline{a}_{x+t:y},$$

which is smaller by amount $(\kappa' - \kappa) \cdot \overline{B} \cdot \overline{a}_{x+t:y}$. The principle of reducing the surrender value paid to the policyholder by the amount of deferred acquisition costs is called *zillmering* or *zillmerization.* In practice, this is taken into account approximatively by reducing from the provision amount

$$Z(t) = \overline{B} \cdot (\kappa' - \kappa) \cdot (m' - t)^{+}.$$

where $m' = \min(m, 10)$. Often in practice contract-by-contract zillmering is not used when calculating technical provisions for annual accounts; instead, an estimate of the sum of individual zillmerings is subtracted from the technical provisions.

The provision of the contract at time t, V_t, is called the *non-forfeiture value*[53] of the contract; the surrender value is $V_t - Z(t)$, i.e. non-forfeiture value less individual zillmering. Technical provisions for annual accounts are calculated as the sum of non-forfeiture values less an estimate of the sum of zillmerings.

2.7.2. *Changes in Contract after Initiation.* In case changes are made in an insurance contract after initiation, they are taken into account by applying the equivalence principle: i.e. surrender value or non-forfeiture value of the contract must remain invariant with respect to the change. A change could be a modification of the insurance cover (e.g. sum insured is changed) or a discontinuation of premium payments.

In case insurance cover is modified but premium payments continue, equivalence principle requires that non-forfeiture value must remain invariant. The equality of non-forfeiture

[52] surrender value = återköpsvärde = takaisinostoarvo

[53] non-forfeiture value = ändringsvärde = muutosarvo

values $V_t^{old} = V_t^{new}$ can be written as

$$S_w^{old} \cdot A_{x+t:w}(\mathbf{V}) + S^{old} \cdot ((1+\phi) \cdot A_{x+t:w} + \epsilon \cdot \overline{a}_{x+t:w}) - (1-\kappa) \cdot \overline{B}^{old} \cdot \overline{a}_{x+t:y} =$$

$$= S_w^{new} \cdot A_{x+t:w}(\mathbf{V}) + S^{new} \cdot ((1+\phi) \cdot A_{x+t:w} + \epsilon \cdot \overline{a}_{x+t:w}) - (1-\kappa) \cdot \overline{B}^{new} \cdot \overline{a}_{x+t:y}.$$

If premium payments are discontinued, then the new premium level is 0 and equivalence principle requires determining the level of insurance cover earned with paid premiums – the *paid-up value* of the contract. Insurance cover may be adjusted either by reducing the sum insured or shortening the contract term. If sums insured are adjusted and the acquisition costs have already been amortized, then the paid-up value is

$$V_t = S_w^{new} \cdot A_{x+t:w}(\mathbf{V}) + S^{new} \cdot ((1+\phi) \cdot A_{x+t:w} + \epsilon \cdot \overline{a}_{x+t:w}).$$

In case deferred acquisition costs are not recoverable during remaining contract term, then we must use surrender value:

$$V_t - Z(t) = S_w^{new} \cdot A_{x+t:w}(\mathbf{V}) + S^{new} \cdot ((1+\phi) \cdot A_{x+t:w} + \epsilon \cdot \overline{a}_{x+t:w}).$$

2.7.3. *Analysis and Surpluses.* Since $\int_a^b V_t' dt = V_b - V_a$, integrating Thiele's equation over $[a,b] \subset [0,n]$ yieds

$$V_b - V_a = \int_a^b \delta \cdot V_t dt + \int_a^b \tau_t \cdot V_t dt + \int_a^b B_t dt - \int_a^b E_t dt - \int_a^b \underline{\nu}_t \cdot \underline{S}_t dt - \int_a^b L_t dt, \tag{182}$$

which states that the increase in insurer's liabilities consists of interest credited to the policies, termination bonus and paid premiums less paid endowment claims, charges for covering the claim costs arising from insured risks and loadings for covering insurer's operating expenses.

Rearranging equation(183) gives

$$V_b - V_a - \int_a^b B_t dt + \int_a^b E_t dt = \int_a^b \delta \cdot V_t dt - \int_a^b (\underline{\nu}_t \cdot \underline{S}_t - \tau_t \cdot V_t) dt - \int_a^b L_t dt. \tag{183}$$

On the right hand side are the assumptions on interest yield, claim costs and expenses used in calculations. Insurers analyse annually how correct these assumptions have been on aggregate (summed over all contracts in the insurer's portfolio). If the investment return from the insurer's investment activity is larger than the assumed interest yield, the insurer has during the year accrued an *interest rate surplus*. Similarly, comparing the second term on the right hand side to paid claim costs less provisions released from terminated contracts yields the insurer's *risk surplus*, and deducting the year's actual operating expenses from loadings term yields the *expense surplus*. Usually there will be surpluses as the law requires the actuarial assumptions used for calculating premiums and provisions (*first order calculation basis*) to be determined prudently to ensure the survival of the insurer and the policyholders' benefits. As the realized experience (*second order basis*) usually turns out to be more advantageous to the insurer than what was assumed when setting premiums, the insurer is on average expected to generate surpluses systematically. A reasonable amount of this surplus should be distributed back to policyholders as *discretionary bonuses*. Analysis of different surpluses can be used to determine to which contracts bonuses are given (e.g. bonuses can be paid to contracts which have generated largest surpluses).

Integration over the whole contract term yields

$$\int_0^n \delta \cdot V_t dt + \int_0^n \tau_t \cdot V_t dt + \int_0^n B_t dt = \int_0^n E_t dt + \int_0^n \underline{\nu}_t \cdot \underline{S}_t dt + \int_0^n L_t dt, \tag{184}$$

which states that the insurer finances its costs from the contract (endowment payments, risk sums and expenses) with its proceeds (premiums, investment income and provisions released due to termination of contract) from the contract.

Example: For an endowment life insurance the change in the technical provision during time interval (a, b) is

$$V_b - V_a - \int_a^b B_t dt = \delta \cdot \int_a^b V_t dt - \int_a^b \mu_{x+t} \cdot (S - V_t) dt - \int_a^b (\kappa \cdot B_t + \phi \cdot \mu_{x+t} \cdot S + \epsilon \cdot S)\, dt.$$

Quantity $S - V_t$ is called *sum at risk at time t*. Most often the time interval considered is the past accounting year, i.e. $[a, b] = [t-1, t]$ and previous equation takes the form

$$\overline{B} + \int_{t-1}^t \delta \cdot V_u du - \Delta V_t = \int_{t-1}^t \mu_{x+u}(S - V_u) du + \left(\kappa \cdot B_t + \phi \cdot S \cdot \int_{t-1}^t \mu_{x+u} du + \epsilon \cdot S\right),$$

i.e. the sum of the insurer's proceeds from risk coverage and expense loadings equal the sum of premium income and technical interest less the increase in liabilities (provision). Considered over the whole contract term, we have

$$\int_0^n B_t dt + \int_0^n \delta \cdot V_t dt = S + \int_0^n \mu_{x+t}(S - V_t) dt + \int_0^n (\kappa \cdot B_t + \phi \cdot S \cdot \mu_{x+t} + \epsilon \cdot S)\, dt.$$

That is, if the realized investment returns, mortality and expenses are equal to the assumptions used in the first order basis, then premiums and investment income are just sufficient to cover claim costs and operating expenses.

3. Multiple State Models in Life Insurance

3.1. **Transition Probabilities.**

We now consider insurance contracts in which the claim payments are conditional on the insured being in a specified *state*, such as e.g. disabled, ill with a specified sickness or dead. For this purpose we assume that an insured person can be in s disjoint (mutually exclusive) states, and may jump from one state to another as time passes.

Definition 3.1.1: Let $\tilde{S}(t)$ be a right-continuous random variable depending on age t of the insured and taking values in $\{1, \ldots, s\}$. Denote by

$$A(y) := \min\left\{t \,\middle|\, \tilde{S}(u) = \tilde{S}(y) \text{ for all } u \in [t, y]\right\},$$

the most recent age of arrival into the state in which the insured is at age y. The probability that an insured person aged y moves from state j to state i by age z is called the *conditional transition probability* and is denoted

$$p_{ij}(z, y \,|\, x) := \mathbb{P}\left(\tilde{S}(z) = i \,|\, \tilde{S}(y) = j,\, A(y) = x\right), \tag{185}$$

for $x \leq y \leq z$. We can gather these transition probabilities into a matrix $p(z, y) = (p_{ij}(z, y \,|\, x))$ containing transition probabilities between all states from age y to age z.

Observe that the previous definition does not exclude the insured visiting other states than i and j in between ages y and z. In the matrix $p(z, y)$, x depends on the initial state j, i.e. is different for different elements of the matrix and is therefore left out of the notation.

We will assume that transition probabilities are differentiable; in practice they may be even discontinuous: if there is a minimum age limit w for old age pension, then the transition probability from active state to pensioner state is zero at any age before age w and strictly positive at age w.

Age x is taken into account in order to be able to have the transition probabilities dependent on for how long the insured has been in state j. For example, the probability of a disabled person becoming fit to work again depends strongly on the duration of the disability. On the other hand, we are assuming that transition probability does not depend on the history of the insured before arriving to state j. In principle this restriction can be overcome by expanding the state space: simply consider being disabled for the first time as a separate state from being disabled for the second time.

3.2. **Markov Chains.**

Definition 3.2.1: The stochastic process $\tilde{S}(t)$ introduced in the previous section is a *Markov chain*, if none of the transition probabilities p_{ij} depend on how long the insured has been in state j. In this case,

$$p_{ij}(z, y) := \mathbb{P}\left(\tilde{S}(z) = i \,|\, \tilde{S}(y) = j\right), \tag{186}$$

for $y \leq z$.

Many processes which are not Markov chains can in theory be transformed to Markov chains by decomposing state j into states indexed by $(j, y-x)$. Since the time spent in state j, $y-x$, is a continuous variable, this leads to infinite (even uncountable) number of states. In principle, this may be solved by discretization (e.g. taking into account only integer part of $y-x$); however, often this leads to unpractically many states.

The transition from state j to state i may happen via some other state k, and the transition probability is the sum of probabilities of all possible transition routes:

Proposition: Let $y \le z$ and $\tilde{S}$ be a Markov chain. For each $u \in [y,z]$, the *Chapman–Kolmogorov equations* hold:

$$p_{ij}(z,y) = \sum_{k=1}^{s} p_{ik}(z,u) \cdot p_{kj}(u,y); \tag{187}$$

in matrix form

$$p(z,y) = p(z,u)p(u,y). \tag{188}$$

Proof: Firstly,

$$\begin{aligned} p_{ij}(z,y) &= \mathbb{P}\left(\tilde{S}(z) = i \,\middle|\, \tilde{S}(y) = j\right) \\ &= \textstyle\sum_{k=1}^{s} \mathbb{P}\left(\tilde{S}(z) = i \,\middle|\, \tilde{S}(u) = k,\, \tilde{S}(y) = j\right) \cdot \mathbb{P}\left(\tilde{S}(u) = k \,\middle|\, \tilde{S}(y) = j\right). \end{aligned}$$

Because $\tilde{S}$ is a Markov chain

$$\begin{aligned} p_{ij}(z,y) &= \textstyle\sum_{k=1}^{s} \mathbb{P}\left(\tilde{S}(z) = i \,\middle|\, \tilde{S}(u) = k\right) \cdot \mathbb{P}\left(\tilde{S}(u) = k \,\middle|\, \tilde{S}(y) = j\right) \\ &= \textstyle\sum_{k=1}^{s} p_{ik}(z,u) \cdot p_{kj}(u,y). \square \end{aligned}$$

Definition 3.2.2: The right derivative of the transition probability

$$\mu_{ij}(y \,|\, x) := \lim_{z \to y+} \frac{p_{ij}(z,y \,|\, x) - p_{ij}(y,y \,|\, x)}{z-y}$$

is called the *transition intensity* from state j to state i. If $i = j$, it is sometimes called *staying intensity*. For a Markov process, we leave out x and $|$. We denote the matrix of transition intensities by $\mu(y)$. If matrix differentiation is defined elementwise, then

$$\frac{d}{dz} p(z,y)\big|_{z=y} = \mu(y). \tag{189}$$

In the following operator $\frac{d}{dz}$ refers to right derivative. Since $1 = \sum_{i=1}^{s} p_{ij}(z,y \,|\, x_j)$, differentiation of both sides with respect to z and setting $z = y$ yields

$$\sum_{i \ne j} \mu_{ij}(y \,|\, x_j) = -\mu_{jj}(y \,|\, x_j), \tag{190}$$

that is, the staying intensity of a state is the negative of the sum of transition intensities to other states.

For $j \neq i$, transition intensity is non-negative and

$$\mu_{ij}(y \mid x) = \lim_{z \to y+} \frac{p_{ij}(z, y \mid x)}{z - y} \geq 0, \tag{191}$$

since $p_{ij}(y, y \mid x) = 0$. Staying intensity is non-positive and

$$\mu_{ii}(y \mid x) = \lim_{z \to y+} \frac{p_{ii}(z, y \mid x) - 1}{z - y} \leq 0, \tag{192}$$

since $p_{ii}(y, y \mid x) = 1$.

Force of mortality is also an intensity, but it was defined in a different way in our previous considerations. However, the definitions are equivalent (i.e. force of mortality is a transition intensity in the sense of this section).

Theorem: Dying is a stochastic process, if we define $\tilde{S}(t) = 1$, if the person is alive at time t, and $\tilde{S}(t) = 2$, if the person is dead at time t. This process is a Markov chain and

$$\mu_{21}(y) = \mu_y,$$

where the left hand side follows the notation of this section and the right hand side follows the notation of previous sections.

Proof: The event of a person alive at age y dying by age z is a Markov process, because probability of death depends only on age z and the fact that the person is alive at age y. Obviously, no additional information is contained in knowing for how long the person has been alive at age y. The transition probability from state 1 to 2 is

$$p_{21}(z, y) = {}_{z-y}q_y = 1 - \exp\left(-\int_0^{z-y} \mu_{y+u} du\right),$$

and hence

$$\mu_{21}(y) = \frac{d}{dz} p_{21}(z, y)\Big|_{z=y} = \left(\mu_{z-y+y} \exp\left(-\int_0^{z-y} \mu_{y+u} du\right)\right)_{z=y} = \mu_y. \ \square$$

Intensity of staying alive is then $-\mu_y$, by virtue of equation (190).

Theorem: Consider ages y and z, $y \leq z$, when $\tilde{S}(t)$ is a Markov chain and E_i is the set of states from which state i is accessible. Then

$$\frac{d}{dz} p_{ij}(z, y) = \sum_{k \in E_i} \mu_{ik}(z) \cdot p_{kj}(z, y), \tag{193}$$

or in matrix form

$$\frac{d}{dz} p(z, y) = \mu(y) p(z, y). \tag{194}$$

Proof: Differentiate Chapman–Kolmogorov equations (187) with respect to z to get

$$\frac{d}{dz} p_{ij}(z, y) = \sum_{k=1}^{s} \frac{d}{dz} p_{ik}(z, u) \cdot p_{kj}(u, y)$$

and set then $u = z$ to get

$$\sum_{k=1}^{s} \frac{d}{dz} p_{ik}(z,u) \cdot p_{kj}(u,y)\big|_{u=z} = \sum_{k=1}^{s} \mu_{ik}(z) \cdot p_{kj}(z,y).$$

The theorem follows, since by definition transition intensities from states $k \notin E_i$ (from which state i is non-accessible) to state i are zero. □

If transition intensities are known, then the transition probabilities can be solved from equation (194) – it is a set of s first order linear differential equations, for which the following general results hold.

3.3. A Very Short Interlude on Linear Differential Equations.

Theorem: Let n be a positive integer and $[a,b] \subset \mathbb{R}$. Consider a set of n linear first order differential equations

$$\begin{cases} y_1'(t) & = c_{11}(t)y_1(t) + \ldots + c_{1n}(t)y_n(t) + f_1(t) \\ \ldots & \ldots \\ y_n'(t) & = c_{n1}(t)y_n(t) + \ldots + c_{nn}(t)y_n(t) + f_n(t), \end{cases}$$

where functions c_{ij} are continuous and functions y_i differentiable on $[a,b]$. This can be written in matrix form as $\underline{y}'(t) = C(t) \cdot \underline{y}(t) + \underline{f}(t)$, where $C = (c_{ij})$, $\underline{y} = (y_1, \ldots, y_n)$ and $\underline{f} = (f_1, \ldots, f_n)$. If the vector of initial values $\underline{y}(a)$ is known, then this is an *initial value problem.* If $f = 0$, then the equations are *homogeneous.* If $det(C(a)) \neq 0$, then the initial value problem has a unique non-zero solution on $[a,b]$.

We do not prove this theorem here, as it is a basic result in the theory of differential equations. A general solution formula in terms of matrix exponential $\exp(C)$ exists, but writing out the solution componentwise leads to infinite series. In some special cases the solution is expressible algebraically in terms of real-valued functions.

Theorem: Equation

$$c_n(t) \cdot y^{(n)}(t) + \ldots + c_1(t) \cdot y'(t) + c_0(t) \cdot y(t) = f(t) \tag{195}$$

is an *nth order linear differential equation*, where f and c_i, $i = 0, \ldots, n$, are continuous functions. If the initial values $y(a)$, $y'(a)$, …, $y^{(n)}(a)$ are known, this is an *initial value problem.* This initial value problem has exactly n linearly indepedent solutions.

We do not prove this theorem here. The proof is based, among other things, on the fact that any nth order linear differential equation can be transformed into a set of first order linear differential equations. If the coefficient functions are constants, we have the following result.

Theorem: Consider an nth order linear homogeneous differential equation with constant coefficients

$$c_n \cdot y^{(n)}(t) + \ldots + c_1 \cdot y'(t) + c_0 \cdot y(t) = 0,\ t \in [a,b]. \tag{196}$$

The *characteristic equation* of this differential equation is

$$c_n \cdot r^n + \ldots + c_1 \cdot r + c_0 = 0, \tag{197}$$

the left hand side of which is an nth order polynomial in r and, as such, has n zeroes in the complex plane $\mathbb{C}$, by the fundamental theorem of algebra. If the characteristic equation has n different roots $r_i \in \mathbb{C}$, $i = 1, \ldots, n$, then all solutions of equation (196) are of form

$$y(t) = C_n \cdot e^{r_n t} + \ldots + C_2 \cdot e^{r_2 t} + C_1 \cdot e^{r_1 t}, \tag{198}$$

where coefficients $C_i \in \mathbb{C}$ are uniquely determined from the initial values of the initial value problem. If the characteristic equation has a kth order root ($2 \leq k \leq n$) r_j, then in the solution (198) the sum $C_j \cdot e^{r_j t} + \ldots + C_j \cdot e^{r_j t}$ is replaced by

$$C_j e^{r_j t} + C_{j+1} \cdot t^1 \cdot e^{r_j t} + \ldots + C_{j+k-1} \cdot t^{k-1} \cdot e^{r_j t}.$$

Functions appearing in the differential equations encountered in applications considered in this presentation do not always satisfy the continuity requirements of previous theorems. However, the functions are piecewise continuous with a finite number of discontinuity points, and the theorems may be applied piecewise to such subintervals of $[a, b]$ over which the functions are continuous. That is, first solve the equation on the first subinterval with initial value equal to the original initial value, then on the second subinterval with initial value equal to the final value of the preceding subinterval plus any possible additional term arising at the discontinuity point, and so on until the whole interval has been covered.

3.4. **Evolution of the Collective of Policyholders.** In this section we consider a closed (i.e. no new entrants are allowed) collective of policyholders, whose initial distribution to states $\{1, \ldots, s\}$ is known. We are interested in the distribution of policyholders to states at some later date.

Definition 3.4.1: Denote by $\tilde{l}_i(y)$ the *number of policyholders aged y in state i*, and the s-vector of these numbers by $\tilde{l}(y) := (\tilde{l}_i(y))_{i=1}^s$.

Theorem: Suppose that the transitions of a policyholder between states are described by a Markov chain, and that the transition probabilities p_{ij} are the same for all policyholders. If the number of policyholders aged y is $l(y)$, then the conditional expected value of the number of policyholders after time t, conditional on $\tilde{l}(y) = l(y)$, is (in matrix form)

$$l(y + t) := p(y + t, y) l(y)$$

or

$$l_i(y + t) := \sum_{j=1}^{s} p_{ij}(y + t, y) \cdot l_j(y).$$

Then the expected value of the number of policyholders aged y in state i changes in time according to

$$\frac{d}{dt}l_i(y+t) = \sum_{j=1}^{s}\sum_{k\in E_i}\mu_{ik}(y+t)p_{kj}(y+t,y)\cdot l_j(y).$$

For $t=0$, we get

$$\frac{d}{dy}l_i(y) = \sum_{j=1}^{s}\mu_{ij}(y)\cdot l_j(y).$$

Proof: By definition of expected value

$$l_i(y+t) = \mathbb{E}\left(\tilde{l}_i(y+t)\,|\,l(y)\right) = \sum_{j=1}^{s}p_{ij}(y+t,y)\cdot l_j(y).$$

Then

$$\frac{d}{dt}l_i(y+t) = \sum_{j=1}^{s}\frac{d}{dt}p_{ij}(y+t,y)\cdot l_j(y) = \sum_{j=1}^{s}\sum_{k\in E_i}\mu_{ik}(y+t)\cdot p_{kj}(y+t,y)\cdot l_j(y).$$

Since $p_{kj}(y,y)$ is zero for $k\neq j$ and one for $k=j$, the last formula follows. □

3.5. **Solutions for Transition Probabilities.**

Definition 3.5.1: We denote by

$$\tilde{B}(y) := \min\left(t\,|\,\tilde{S}(t)\neq\tilde{S}(y), t>y\right)$$

the *age of departure* from the state the policyholder is in at age y. Probability of a transition in one step is

$$\overline{p}_{ij}(z,y\,|\,x) = \mathbb{P}\left(\tilde{S}(z)=i\,|\,\tilde{S}(u)=i,\ \tilde{B}(y)\leq u\leq z \text{ and } A(y)=x\right).$$

Probability of no transition is

$$\overline{p}_{jj}(z,y\,|\,x) = \mathbb{P}\left(\tilde{S}(t)=j,\ t\in[x,z]\right).$$

We can rewrite equation (193) as

$$\frac{d}{dz}p_{ij}(z,y) = \mu_{ii}(z)\cdot p_{ij}(z,y) + \sum_{k\neq i,k\in E_i}\mu_{ik}(z)\cdot p_{kj}(z,y), \tag{199}$$

which is a first order linear differential equation for functions p_{ij} and can be integrated to form (using multiplication with integrating factor $\exp\left(-\int\mu_{ii}(u)du\right)$)

(200)

$$p_{ij}(z,y) = p_{ij}(y,y)\cdot\exp\left(\int_y^z\mu_{ii}(u)du\right)+\int_y^z\left(\exp\left(\int_u^z\mu_{ii}(v)dv\right)\cdot\sum_{k\neq i,k\in E_i}\mu_{ik}(u)\cdot p_{kj}(u,y)\right)du.$$

For $i \neq j$, we have $p_{ij}(y,y) = 0$, and hence in that case the above general formula simplifies to

$$p_{ij}(z,y) = \int_y^z \left(\exp\left(\int_u^z \mu_{ii}(v)dv \right) \cdot \sum_{k\neq i, k\in E_i} \mu_{ik}(u) \cdot p_{kj}(u,y) \right) du, \tag{201}$$

which states that the probability of a policyholder who at age y is in state j to be in state i at age z is the sum of probabilities of a policyholder who at age y is in state j being in state k at age u, in which age the policyholder moves to state i where he/she stays until age z with probability $\exp(\int_u^z \mu_{ii}(v)dv) < 1$ (since $\mu_{ii} < 0$).

For $i = j$, we have $p_{ij}(y,y) = 1$, and hence in that case the above general formula simplifies to

$$p_{jj}(z,y) = \exp\left(\int_y^z \mu_{jj}(u)du \right) + \int_y^z \left(\exp\left(\int_u^z \mu_{jj}(v)dv \right) \cdot \sum_{k\neq j, k\in E_i} \mu_{jk}(u) \cdot p_{kj}(u,y) \right) du. \tag{202}$$

which states that the probability that a policyholder who at age y is in state j is also at age z in state j is equal to the sum of the probability that the policyholder stays in state j all the time from age y to age z and the probability that the policyholder visits other states $k \neq j$ in between but returns to state j by age z.

Previous two transition formulas hold true only for Markov chains. The following result, in contrast, holds more generally.

Theorem: The probability of a policyholder staying in state j all the time from age y to age z, when he/she has been in state j since age x, is

$$\overline{p}_{jj}(z,y \,|\, x) = \exp\left(\int_y^z \mu_{jj}(u \,|\, x)du \right) = \Pi_{k\neq j} \exp\left(- \int_y^z \mu_{kj}(u \,|\, x)du \right). \tag{203}$$

The probability of a one-step transition from state j to state i is

$$\begin{aligned} \overline{p}_{ij}(z,y \,|\, x) &= \textstyle\int_y^z \left(\overline{p}_{ii}(z,u \,|\, u) \cdot \mu_{ij}(u \,|\, x) \cdot \overline{p}_{jj}(u,y \,|\, x) \right) du \\ &= \textstyle\int_y^z \left(\exp\left(\int_u^z \mu_{ii}(v \,|\, u)dv \right) \cdot \mu_{ij}(u \,|\, x) \cdot \exp\left(\int_y^u \mu_{jj}(v \,|\, x)dv \right) \right) du. \end{aligned} \tag{204}$$

Proof: For the first equation, observe that since there is no transition to other state,

$$\frac{d}{dz} \overline{p}_{jj}(z,y \,|\, x) = \mu_{jj}(z \,|\, x) \cdot \overline{p}_{jj}(z,y \,|\, x). \tag{205}$$

(The reasoning goes as follows: the first equation we encountered in the proof of Chapman–Kolmogorov equations (187) is true for a process $\tilde{S}$ which is not necessarily a Markov chain, and since we are now dealing with a staying intensity, in that equation we now have $i = j = k$ and the equation reduces to

$$p_{jj}(z,y\,|\,x) =$$

$$= \mathbb{P}\left(\tilde{S}(z) = j\,|\,\tilde{S}(u) = j, \tilde{S}(y) = j, A(y) = x\right) \cdot \mathbb{P}\left(\tilde{S}(u) = j\,|\,\tilde{S}(y) = j, A(y) = x\right);$$

but since now $A(y) = A(u)$ as in both ages we have been uninterrupted in state j from age x, this can be written as

$$\begin{aligned} p_{jj}(z,y\,|\,x) &= \mathbb{P}\left(\tilde{S}(z) = j\,|\,\tilde{S}(u) = j, A(u) = x\right) \cdot \mathbb{P}\left(\tilde{S}(u) = j\,|\,\tilde{S}(y) = j, A(y) = x\right) \\ &= p_{jj}(z,u\,|\,x) \cdot p_{jj}(u,y\,|\,x); \end{aligned}$$

differentiation with respect to z and setting then $u = z$ yields the differential equation (205).) Solution of this differential equation with initial condition $\overline{p}_{jj}(y,y\,|\,x) = 1$ is the expression after first equality sign in the first equation of the theorem. The expression after second equality follows from the fact that staying intensity is the negative of the sum of transition intensities.

If the transition can only occur from state j to state i, then for a Markov chain

$$\overline{p}_{ij}(z,y) = \int_y^z \left(\exp\left(\int_u^z \mu_{ii}(v)dv\right) \cdot \mu_{ij}(u) \cdot \overline{p}_{ij}(u,y)\right) du,$$

i.e. after being in state j at age y the policyholder moves at age u from state j to state i where he/she stays until age z. If the policyholder has been in state j since age x, then according to preceding interpretation the formula takes the form

$$\overline{p}_{ij}(z,y\,|\,x) = \int_y^z \left(\exp\left(\int_u^z \mu_{ii}(v\,|\,u)dv\right) \cdot \mu_{ij}(u\,|\,u) \cdot \overline{p}_{ij}(u,y\,|\,x)\right) du,$$

and this is true even if we are not dealing with a Markov chain, by a similar argument used in deriving the differential equation (205). Inserting into this equation the expressions for $\overline{p}_{ii}(z,u\,|\,u)$ and $\overline{p}_{jj}(u,y\,|\,x)$ from the first equation of the theorem we obtain the second equation of the theorem. □

Example: Consider a Markov chain with an infinite (but countable) number of states $\{k, 1, 2, 3, \ldots\}$ such that transition intensities from state i to state $i+1$ are ν_i, transition intensities from state i to state k are μ_i and all other transition intensities are zero. The model states could be interpreted as follows: k is"dead", odd i's are "active after being disabled for $i-1$ times" and even i is "disabled for $(i-1)$'th time". Suppose that a policyholder is in state 1 at age x and denote

$$p_{j1}(t) := \mathbb{P}\left(\tilde{S}(x+t) = j\,|\,\tilde{S}(x) = 1\right),$$

and denote the mortality intensity of a policyholder in state i by $\mu_i(t) := \mu_{ki}(x+t)$ and the transition intensity $\nu_i(t) := \mu_{i+1,i}(x+t)$. Since there is no returning to earlier states, $\mu_{ij} = 0$ for $j > i$. Then, by the results of this chapter,

$$\frac{d}{dt}p_{11}(t) = -(\mu_1(t) + \nu_1(t)) \cdot p_{11}(t).$$

Integrating this equation with initial condition $p_{11}(0) = 1$ yields

$$p_{11}(t) = \exp\left(-\int_0^t (\mu_1(u) + \nu_1(u))du\right).$$

Since the only states from which j is accessible are $j-1$ and j (i.e. $E_j = \{j-1, j\}$),

$$\frac{d}{dt}p_{j1}(t) = \nu_{j-1}(t) \cdot p_{j-1,1}(t) - (\mu_j(t) + \nu_j(t)) \cdot p_{j1}(t),$$

for $j = 2, 3, \ldots$. Taking into account initial condition $p_j(0) = 0$ yields

$$p_{j1}(t) = \int_0^t \exp\left(-\int_u^t (\mu_j(z) + \nu_j(z))\right) \cdot \nu_{j-1}(u) \cdot p_{j-1,j}(u)du.$$

From this probabilities of being in state j can be calculated recursively. Example generalizes in a straightforward fashion to the case where transitions from state i to any state $j > i$ occur with intensity ν_{ij}.

3.6. **An Example of a Disability Model.** Consider a three-state disability insurance with states a: "active", i: "invalid" and k: "dead". Disability pension is paid while the policyholder is invalid. Transitions may occur from state a to states i or k and from state i to states a or k. From state k no transitions back to other states may occur. If we assume that transitions only depend on the age of the policyholder (i.e. that we are dealing with a Markov chain), then transition probabilities can be derived from corresponding intensities. If the policyholder is aged x at the beginning, then transition intensities at age $x + t$ are denoted

i: $\tau(t) :=$ intensity of becoming invalid;
ii: $\nu(t) :=$ intensity of returning to active state;
iii: $\mu_a(t) :=$ mortality of an active person;
iv: $\mu_i(t) :=$ mortality of an invalid.

Let the number of policyholders aged x at time $t = 0$ be $l(0) = l_a(0) + l_i(0)$, where $l_a(0)$ is the number of active policyholders and $l_i(0)$ is the number of invalid policyholders. We obtain a set of linear differential equations

$$l'_j(t) = \sum_s \mu_{js}(t) \cdot l_s(t),$$

which now takes the form

$$\begin{cases} l'_a(t) &= -(\mu_a(t) + \tau(t)) \cdot l_a(t) + \nu(t) \cdot l_i(t) \\ l'_i(t) &= -(\mu_i(t) + \nu(t)) \cdot l_i(t) + \tau(t) \cdot l_a(t.) \end{cases} \tag{206}$$

The solution of this pair of equations is in general not expressible in terms of finite sums. However, as was stated in the section on linear differential equations, subject to some regularity conditions the pair of equations has a unique solution (important for the use of numerical solution methods to make sense). We will next derive the transition probabilities, expected present values of life annuities for disability pensions and the net premium for disability pension assuming that the solution is known. After that, we look at under which conditions the pair of equations is in practice solvable.

Proposition: Let all policyholders be in active state at time $t = 0$, i.e. $l_a(0) > 0$ and $l_i(0) = l_k(0) = 0$. At time t

$$\begin{aligned} p_{aa}(x+t,x) &= \mathbb{P}\left(\tilde{S}(x+t) = a \,|\, \tilde{S}(x) = a\right) = \tfrac{l_a(t)}{l_a(0)} \\ p_{ia}(x+t,x) &= \mathbb{P}\left(\tilde{S}(x+t) = i \,|\, \tilde{S}(x) = a\right) = \tfrac{l_i(t)}{l_a(0)} \\ p_{ka}(x+t,x) &= \tfrac{l_k(t)}{l_a(0)} = 1 - p_{aa}(x+t,x) - p_{ia}(x+t,x). \end{aligned} \tag{207}$$

Proof: Formula $l_{\cdot}(y+t) = \sum_{j=1}^{s} p_{\cdot,j}(y+t,y) \cdot l_j(y)$ reduces in this case to $l_a(t) = p_{aa}(x+t,x) \cdot l_a(0)$, since $l_i(0) = l_k(0) = 0$. This proves the first equation. The second one is proved similarly, and the third follows from the preceding two. □

Proposition: The policyholder has obtained insurance while in active state at age x. The insurer has agreed to pay a continuous unit disability pension for the duration of disability at most until age $w = x + n$. Premiums are paid continuously while the policyholder is in active state, however at most for m years. Force of interest is δ.

i: The expected present value of the future disability pension for a policyholder active at age $x + t$ is

$$\overline{a}^{ia}_{x+t:w} := \int_t^n p_{ia}(x+u, x+t) \cdot e^{-\delta(u-t)} du. \tag{208}$$

ii: The expected present value of the sum of ongoing and future disability pensions for a policyholder invalid at age $x + t$ is

$$\overline{a}^{ii}_{x+t:w} := \int_t^n p_{ii}(x+u, x+t) \cdot e^{-\delta(u-t)} du. \tag{209}$$

iii: The expected present value of the ongoing disability pension for a policyholder invalid at age $x + t$ is

$$\overline{a}^{i}_{x+t:w} := \int_t^n \overline{p}_{ii}(x+u, x+t) \cdot e^{-\delta(u-t)} du = \int_t^n \exp\left(-\int_t^u (\delta + \mu_i(s) + \nu(s)) ds\right) du. \tag{210}$$

iv: The continuous level insurance premium is

$$\overline{P} = \overline{E} \cdot \frac{\overline{a}^{ia}_{x:\overline{n}|}}{\overline{a}^{aa}_{x:\overline{m}|}}, \tag{211}$$

where $\overline{E}$ is the annual amount of disability pension and

$$\overline{a}^{aa}_{x:\overline{m}|} = \int_0^m p_{aa}(x+u, x+t) e^{-\delta \cdot u} du$$

is a life annuity payable while in active state.

Proof: Denote the indicator of being invalid by $\tilde{J}_{x+t}(u) := I_i(\tilde{S}(x+u))$, where I_i is the indicator of state i (invalid). Then the expected value of the random present value of a continuous n-year unit pension is

$$\overline{a}^{ia}_{x+t:w} = \mathbb{E}\left[\int_t^n \tilde{J}_{x+t}(u) \cdot e^{-\delta(u-t)} du\right] = \int_t^n \mathbb{E}\left[\tilde{J}_{x+t}(u)\right] \cdot e^{-\delta(u-t)} du.$$

Part (i) of the proposition follows immediately, since

$$\mathbb{E}\left[\tilde{J}_{x+t}(u)\right] = \mathbb{E}\left[I_i(\tilde{S}(x+u))\right] = \mathbb{P}\left(\tilde{S}(x+u) = i \,|\, \tilde{S}(x+t) = a\right) = p_{ia}(x+u, x+t).$$

Parts (ii) and (iii) are proved in completely similar fashion; for the second equality in (iii), equation (203) is used. Part (iv) follows from equivalence principle. □

The insurer's technical provision for an invalid policyholder is the sum of expected present values of ongoing and future disability pensions. The expected present value of ongoing disability pension is a part of the outstanding claim provision, while the expected present value of future disability pensions is a part of mathematical provision.

We conclude this example by considering how to solve the number of policyholders in different states at time t (i.e. equations (206)). As previously stated, in general a numerical solution method is called for. However, if all the transition intensities are constants, a closed form analytical solution can be obtained. In this case, differentiating the first equation in (206) yields

$$l''_a(t) = -(\mu_a + \tau) \cdot l'_a(t) + \nu \cdot l'_i(t).$$

Inserting into this equation the expression for derivative l'_i from the second equation of (206) yields

$$\begin{aligned} l''_a(t) &= -(\mu_a + \tau) \cdot l'_a(t) + \nu \cdot [-(\mu_i + \nu) \cdot l_i(t) + \tau \cdot l_a(t)] \\ &= -(\mu_a + \tau) \cdot l'_a(t) - \nu \cdot l_i(t) \cdot (\mu_i + \nu) + \nu \cdot \tau \cdot l_a(t) \end{aligned}$$

Inserting into this an expression for $\nu \cdot l_i(t)$ solved from the first equation of (206) yields a second order linear homogeneous differential equation with constant coefficients

$$l''_a + A \cdot l'_a + B \cdot l_a = 0, \tag{212}$$

where $A = \tau + \nu + \mu_a + \mu_i$ and $B = (\mu_a + \tau) \cdot (\mu_i + \nu) - \tau \cdot \nu$. This is solvable using the characteristic equation $r^2 + A \cdot r + B = 0$. The general solution is

$$l_a(t) = C_1 \cdot e^{r_1 t} + C_2 \cdot e^{r_2 t},$$

where r_1 and r_2 are the two different roots of the characteristic equation. Inserting $l_a(t)$ into the second equation of (206) yields a first order differential equation for $l_i(t)$, which can be solved to yield

$$l_i(t) = \exp\left(-\int_0^t (\mu_i+\nu)du\right)\cdot\left(l_i(0)+\int_0^t \tau\cdot(C_1\cdot e^{r_1 u}+C_2\cdot e^{r_2 u})\cdot\exp\left(\int_0^u (\mu_i+\nu)ds\right)du\right).$$

For disability pensions, constant transition intensities are not really compatible with reality. Another assumption leading to closed form solution in this example is assuming that $\mu_i(t) = \mu_a(t) =: \mu(t)$, i.e. that the mortalities of active and disabled policyholders are the same. This is also typically not quite realistic, as most often the mortality of invalids is higher.

4. On Asset–Liability Management

Asset–Liability Management (ALM) refers to a holistic, total balance sheet approach in managing an insurer's assets and liabilities as a whole. This means integrating the work of *actuarial department* focusing on the measurement, analysis and projection of insurer's liabilities and their risks and the work of *investment department* focusing on the risks and rewards of asset classes and asset strategies. In particular, ALM is about recognizing the importance of investment activities being in accordance with requirements set by the nature of the liabilities. One way of looking at ALM is to consider it as an optimization problem where our goal is to maximize financial returns subject to the constraint that we must be able to cover our liabilities (with sufficiently high probability).

Classical concepts related to ALM are *matching* and *immunization*. Matching of assets with liabilities involves in its purest form structuring the flow of asset proceeds to coincide exactly with the outgo with respect to liabilities. Doing this perfectly is of course possible only in a deterministic world where no uncertainty is associated with future cash flows. Depending on the nature of the liability, a close matching, however, may be possible – some life insurance products such as guaranteed income bonds can be closely matched with investments made in government fixed income securities. In contrast, matching assets may be exceedingly hard to find for e.g. some non-life insurance products. Immunization (first proposed by Redington in 1952) generalizes the concept of matching by requiring investing in such a way that a change in the value of assets is exactly matched by a change in the value of liabilities, i.e. that liabilities are immunized with respect to changes in value due to changes in some underlying factors (in Redington's model, interest rates).

4.1. **Classical Asset-Liability Theory.** Suppose that insurer's expected net liabilities at time t are given by L_t, and that the insurer's expected asset proceeds payable at time t are given by A_t. Then with a force of interest δ, the present value of expected future liabilities is

$$L(\delta) := \int_0^\infty L_z e^{-\delta z} dz$$

and the present value of expected future asset proceeds is

$$A(\delta) := \int_0^\infty A_z e^{-\delta z} dz.$$

Suppose now that assets and liabilities are initially matched under interest rate assumption δ so that $A(\delta) = L(\delta)$, and the interest changes from δ to $\delta + \epsilon$. Both L and A will now change, and by Taylor approximation

$$L(\delta + \epsilon) \approx L(\delta) + \epsilon \cdot L'(\delta) + \frac{\epsilon^2}{2} \cdot L''(\delta)$$

and

$$A(\delta + \epsilon) \approx A(\delta) + \epsilon \cdot A'(\delta) + \frac{\epsilon^2}{2} \cdot A''(\delta).$$

Hence

$$A(\delta+\epsilon)-L(\delta+\epsilon)\approx A(\delta)-L(\delta)+\epsilon\cdot\left(A'(\delta)-L'(\delta)\right)+\frac{\epsilon^2}{2}\cdot\left(A''(\delta)-L''(\delta)\right).$$

The first term of the right hand side of above equation is zero since assets and liabilities were initially matched. If the change in interest rate should cause no profit or loss, then obviously all successive derivatives should be equal to zero. For small changes the first derivative is the most important, and a portfolio with $A'(\delta)-L'(\delta)=0$ is called an immunized portfolio. If furthermore $A''(\delta)-L''(\delta)\geq 0$, then any sufficiently small change in interest rate will result in a profit. Based on these observations, a satisfactory immunization strategy can be characterized by $A'(\delta)-L'(\delta)=0$ and $A''(\delta)-L''(\delta)\geq 0$.

While this seems to make things very simple, it should be observed that in practice: the previous two equations have usually an infinite number of solutions; the equations define the position at a moment in time – their solutions change continuously; solution is dependent on the interest rate (which is what interest rate, exactly?); the solution is based on an approximation argument valid for small changes only; ...and many more problems and complications can be found.

4.2. **Immunization of Liabilities with Bonds.** A very natural idea is to use default-free bonds (whose price or value is a function of interest rates) to construct an immunization strategy against interest rate changes. We will consider three strategies: multi-period immunization, generalized immunization and cash flow matching.

We need to be able to compute the yield and duration of a bond portfolio; while in principle this can be done by considering the portfolio as a single bond and solving the yield as we did for single bonds, in practice this is thought to be too cumbersome and approximations are used. The yield of a bond portfolio is approximated by the present value weighted average of individual bond yields

$$y_p\approx\frac{\sum_{k=1}^N P_k y_k}{\sum_{k=1}^N P_k},$$

and similarly the bond portfolio's duration is approximated by the present value weighted average of individual bond durations

$$D_p\approx\frac{\sum_{k=1}^N P_k D_k}{\sum_{k=1}^N P_k}.$$

When immunizing liabilities, we will employ a generalization of (Macaulay) duration, so-called *ith generalized duration* which is defined for an individual bond k as

$$D^i_{Gen:k}:=\sqrt{W^i_k},$$

where

$$W^i_k:=\frac{1}{P_k}\sum_{t=0}^{n-1}\frac{((\tau+t/m)^i CF_{\tau+t/m}}{(1+y_k)^{\tau+t/m}}$$

(observe that $D^1_{Gen:k} = D_{Mac:k}$). The generalized duration of a bond portfolio is approximated by

$$D^i_{Gen:p} \approx \frac{\sum_{k=1}^{N} P_k D^i_{Gen:k}}{\sum_{k=1}^{N} P_k}.$$

Consider now an insurance product in which the insurer has promised a guaranteed rate of return to premium, i.e. the policyholder or beneficiaries receive a specified guaranteed payment at a specified future date. We assume that the insurer invests in a single bond (using previous approximations the considerations generalize to a bond portfolio). Let h denote the investment horizon and n_1 the number of coupons that will be paid during investment horizon. The value of the reinvested coupons on the horizon date (at time h) is

$$I = \sum_{t=0}^{n_1-1} CF_{\tau+t/m} \cdot (1+y)^{h-\tau-t/m},$$

and the present value of the bond at horizon date is

$$H = \sum_{t=n_1}^{n-1} \frac{CF_{\tau+t/m}}{(1+y)^{\tau+t/m-h}}.$$

The final wealth at the horizon date is $W = I + H$, and an increase in yield will result in a higher I value and a lower H value. The derivative of the final wealth at the horizon date with respect to the yield is

$$\begin{aligned}
\frac{\partial W}{\partial y} &= \frac{\partial}{\partial y}\left[\sum_{t=0}^{n_1-1} CF_{\tau+t/m} \cdot (1+y)^{h-\tau-t/m} + \sum_{t=n_1}^{n-1} \frac{CF_{\tau+t/m}}{(1+y)^{\tau+t/m-h}}\right] \\
&= \frac{h}{(1+y)^{1-h}}\left(\sum_{t=0}^{n_1-1} \frac{CF_{\tau+t/m}}{(1+y)^{\tau+t/m}} + \sum_{t=n_1}^{n-1} \frac{CF_{\tau+t/m}}{(1+y)^{\tau+t/m}}\right) - \\
&- \frac{1}{(1+y)^{1-h}}\left(\sum_{t=0}^{n_1-1} \frac{(\tau+t/m)CF_{\tau+t/m}}{(1+y)^{\tau+t/m}} + \sum_{t=n_1}^{n-1} \frac{(\tau+t/m)CF_{\tau+t/m}}{(1+y)^{\tau+t/m}}\right) \\
&= P \cdot (1+y)^{h-1} \cdot (h - D_{Mac})
\end{aligned}$$

This implies that if

$$h - D_{Mac} \begin{cases} < 0 \Rightarrow \frac{\partial W}{\partial y} < 0 \Rightarrow & \text{market risk} \\ = 0 \Rightarrow \frac{\partial W}{\partial y} = 0 \Rightarrow & \text{immunized ALM portfolio} \\ > 0 \Rightarrow \frac{\partial W}{\partial y} > 0 \Rightarrow & \text{reinvestment risk} \end{cases}$$

Observe that previous analysis is based on a parallel shift of the yield curve. To summarize, for a bond portfolio a fall in interest rates will cause the portfolio to increase in value but the coupons will be reinvested at a lower rate, while a hike in interest rates will cause the portfolio to decrease in value but the coupons will be reinvested at a higher rate. At portfolio's Macaulay duration the two opposite effects will approximately offset each other. To immunize a single liability against yield changes we need to choose a bond portfolio such that

(1) Macaulay duration is equal to the investment horizon;
(2) initial present value of the cash flows equals the present value of future liability.

However, immunization is only achieved if yield changes are parallel; moreover, as yield changes, the duration changes too. Continued immunization hence requires rebalancing the portfolio as time passes, and even then immunization is with respect to parallel changes in the yield curve only.

Obviously a large number of portfolios fulfill the immunization requirements above. However, the more dispersion the cash flow of the portfolio has, the more immunization risk it has. The dispersion of the cash flow of the bond portfolio with respect to the investment horizon h can be measured with

$$\frac{\sum_{t=0}^{n-1}(\tau+t/m-h)^2\frac{CF_{\tau+t/m}}{(1+y)^{\tau+t/m}}}{\sum_{t=0}^{n-1}\frac{CF_{\tau+t/m}}{(1+y)^{\tau+t/m}}}.$$

Hence we might choose the immunized portfolio with the smallest dispersion among the possible alternatives.

In reality, we usually have multi-period liabilities. In this case we can decompose the liabilities to single period liabilities each of which is immunized by a subportfolio of bonds. A multi-period immunization strategy must satisfy the following conditions:

(1) the Macaulay duration of the total bond portfolio must equal the duration of the portfolio of liabilities;
(2) the Macaulay duration distribution of the bond subportfolios must have a wider range than the duration distribution of the liabilities;
(3) the present value of the cash flow from a subportfolio must equal the present value of the corresponding liability stream; and
(4) for each liability the chosen subportfolio has the smallest dispersion among available alternatives.

In generalized immunization, the bond portfolio is chosen in such a way that its generalized durations up to some order match (immunize) the generalized durations of the liability stream. In practical applications the order is rarely higher than 1, which corresponds to usual immunization (matching of Macaulay durations).

In cash flow matching, the last (in time) individual liability cash flow is matched by investment in a bond with maturity equal to the time at which the liability is due and principal equal to the nominal value of the liability. After this, the remaining liability cash flows are reduced by the amount of coupon payments that will be received from the first bond, and then the next to last liability is matched similarly with a bond. The process continues until all liabilities are matched. Observe that a perfectly cash flow matched portfolio is necessarily duration matched, but there may be many duration matched portfolios which are not cash flow matched. The benefits of this approach are that there is no need to compute any durations, no rebalancing is needed and there is no risk that the liabilities cannot be satisfied (unless a bond issuer defaults). The downside of cash flow matching is its high cost: finding matching bonds can be burdensome and still the matching cannot usually be perfect, so that more capital than necessary will need to be set aside to match the liabilities – this may be inefficient capital management, because instruments used for matching are typically low risk and low return bonds and the insurer is not able to take advantage of high return investment opportunities. Naturally, one way of looking at this is that the purpose of a matching portfolio is first and foremost to match the liabilities, while the free reserves (i.e. excess of assets over and above the liabilities) can be used to generate high returns. An investment approach called *core–satellite* is related to this idea: the core portfolio attempts to hedge or replicate the liabilities, while the satellite portfolios concentrate on achieving high returns.

4.3. **Stochastic Asset–Liability Management.** ALM is concerned with risks and rewards of investment strategies and insurance liabilities as a whole. First models of this type were *static*, *single-period* models where a closed form solution was often obtainable. However, to take properly into account the complex time-dependent interrelationships and dependencies of assets and liabilities, and the impact of future management actions, a *dynamic* approach using *stochastic simulations* is called for. Sufficiently realistic models will almost invariably be too complex to be analytically solvable in closed form. However, they can be used to simulate the development of assets and liabilities. The huge advances in computing power have made this a more and more practical solution.

Competitive pressures have added complexity to the products offered by insurance companies: many contracts have implicit or explicit *guarantees* and *embedded options*, which are difficult or impossible to evaluate otherwise than by simulations. Even if the risks are theoretically hedgeable, the hedging may in practice be very costly or impossible to implement and the insurer may be willing or forced to leave part of the risk unhedged. Typical questions of interest are

- **i:** What future economic conditions might lead to problems (losses) and how significant is the impact of the risk if realized?
- **ii:** Are the asset allocation, business strategy and bonus policy appropriate? What actions could be taken to mitigate the risks?
- **iii:** How much capital is required to support the mismatch of assets and liabilities and what is the trade-off between risk and return of this capital?

Future developments may be described by different *economic scenarios*. The cash flows of a contract can then be evaluated under different scenarios to obtain an understanding of how the value of the contract is impacted by different future developments. To achieve this, historical or deterministic pre-specified scenarios can also be used, which reduces the computational effort; however, using a large number of stochastically simulated scenarios has the advantage of giving a more complete picture of the future possibilities, including the impact of rare events and dependencies which may not have historical precedents in available data and also may not be easily covered by a set of deterministic scenarios constructed based on expert judgment (it is easy to overlook some less obvious possibilities). Of course, this requires that the stochastic dynamics are appropriately specified; however, despite this being a non-trivial task, building scenario-generating stochastic models with reasonably realistic assumptions is likely to provide a more complete picture of risks and rewards of assets and liabilities which include complex insurance products and financial instruments than just relying on a few deterministic scenarios.

In an ALM model a special module often called *Economic Scenario Generator (ESG)* usually generates the scenarios for economic variables. Based on input from actuarial department and from investment department, assets and liabilities are projected through each scenario to obtain a range of possible outcomes. Simulation results can then be analyzed with a view to the objectives of risk and reward. Especially with longer planning horizons, it is important to incorporate future management actions and insurer's business and investment strategy into modeling. The longer the horizon, the more untenable is an assumption of a fixed investment strategy. An investor will react to changes in the economic environment and this should be reflected in a long term modeling of assets and liabilities.

Broadly speaking, the purpose of stochastic asset simulation can be *prediction*, *pricing* or *risk analysis*. Prediction is usually very short-term, as markets adapt to changing conditions and prediction of market movements on longer term is quite difficult. Pricing applications often employ absence of arbitrage conditions to check that prices are consistent with the prices of traded securities. Risk analysis evaluates the potential rewards and risks of various investment and liability management strategies, to facilitate avoiding undesired (unsustainable) risks and accepting some other risks to create opportunities for enhanced returns.

Common stochastic asset modeling approaches in insurance are

- **i:** *mean–covariance approach*, which relies on specifying expected values, variances and covariances of assets – obviously, this approach traditionally relies on normal or lognormal time-independent distribution assumptions and hence is often lacking in realism;
- **ii:** *autoregressive approach with a cascade structure*, where variables have hierarchical relationships which are described by autoregressive processes – a well-known example is the *Wilkie model*;
- **iii:** *a stochastic differential equation approach with a cascade structure*, where variables have hierarchical relationships which are described by stochastic differential equations – an example is so-called *Towers Perrin global capital market scenario generating system* or *CAP-link system.*

Stochastic asset models can be divided to *equilibrium models*, which attempt to reflect the long term average behavior of an economy in equilibrium, and *arbitrage-free models*, which are calibrated to reproduce certain market prices exactly (and may then be used to derive market-consistent prices for other assets). Equilibrium models do not usually fit well to observed market data, as variables are rarely at their long term average values, and their calibration is difficult as historical data (even if it covers sufficiently long period) may contain structural breaks, regime shifts and unusual features unlikely to be repeated. Wilkie model is an example of an equilibrium model, where the fundamental economic variable is price inflation, which influences the yields of assets, which in turn influence asset prices. Arbitrage-free models can be fitted exactly to current market prices and hence are well suited for determination of market consistent prices, but as they are usually based on risk neutral valuation, they are not really suitable for realistic projection of the future (as the real world is not risk neutral).

4.4. **On Market-Consistent Valuation and Stochastic Discounting.** Since by definition ALM deals with assets and liabilities in an integrated manner, it is important that assets and liabilities are measured consistently with each other. This may be rephrased as follows: the applied *accounting principles* should be the same for both sides of the balance sheet. Traditionally this has not been the case, as asset values are usually measured at *fair value* or *market value*, that is, at a price with which the asset could be exchanged in a voluntary transaction between two knowledgeable parties (*economic value*). It is important to realize that market prices *per se* have no absolute meaning – in some cases other valuation principles might be more appropriate – but the point in using market prices is precisely the above mentioned switching property:market price is the price at which the asset or liability could be transferred from its owner to a new owner in a fair transaction. In contrast, liabilities have been measured using *actuarial valuation* which determines the value of a liability as a sum of *best estimate* (i.e. expected value) of its cash flows plus a *risk margin* or *prudential margin* over and above the best estimate, to cover the risks associated with uncertainty in the best estimate. In a market-consistent balance sheet, both assets and liabilities should be valued at market value. However, determination of market price is not always straightforward: while for an asset with active deep and liquid markets, the observed market price can be used, most insurance liabilities do not have an active market at all or the existing market is not deep and liquid.

An important point to observe is that in reality, interest rates are stochastic. Hence the traditional deterministic discount rate used in classical actuarial mathematics when valuing liabilities is not consistent with either reality (where we are, in a sense, dealing with the physical or real world probability measure $\mathbb{P}$), or with market-consistent valuation in accordance with the principles of mathematical finance (where we are dealing with the risk neutral probability measure $\mathbb{Q}$). Things get more interesting (or, alternatively, difficult) with stochastic interest rates, as then, for example, with $\tilde{r}$ a stochastic interest rate,

$$1 = \mathbb{E}\left(\frac{1+\tilde{r}}{1+\tilde{r}}\right) \neq \mathbb{E}(1+\tilde{r}) \cdot \mathbb{E}\left(\frac{1}{1+\tilde{r}}\right) > 1$$

(since $\mathbb{E}(\tilde{X}\tilde{Y}) = Cov(\tilde{X},\tilde{Y}) + \mathbb{E}(\tilde{X}) \cdot \mathbb{E}(\tilde{Y})$ and for choices $\tilde{X} = 1+\tilde{r}$, $\tilde{Y} = 1/\tilde{X}$ the covariance is negative). So we cannot exchange stochastic discounting and stochastic accumulation with each other. In calculations of present values of stochastic cash flows, we need to use *stochastic discount factors* which are called *deflators* (or *state price densities* by financial economists). A deflator $\tilde{\phi}_t$ transports cash amount $\tilde{X}_t$ at time t to time 0. This is a stochastic transportation, i.e. the cash flow need not be independent of the deflator, and hence the present value

$$V(\tilde{X}_t) = \mathbb{E}[\tilde{\phi}_t \cdot \tilde{X}_t] \neq \mathbb{E}[\tilde{\phi}_t] \cdot \mathbb{E}[\tilde{X}_t].$$

Many life insurance contracts contain financial options which depend on economic or financial variables influencing the future evolution of the deflator, and hence we need to work with deflators to obtain a realistic model taking into account the interdependencies between variables. That is, deflators allow modeling of options and guarantees in insurance policies. Observe that here the physical probability measure is used – using the risk neutral probability measure, we obtain a state independent discount factor, which simplifies calculations but then the information on real world dependencies is lost and with that the possibilities for realistic modeling of options and guarantees.

From an insurer's point of view the risks can be divided into *financial risks* and *technical risks*. Financial risks are marketable risks which can be traded in an active market, such as e.g. interest rate risk and stock market risk. Technical risks are traditionally non-marketable, non-traded risks such as e.g. longevity, mortality, morbidity and natural catastrophe risk. Observe that many traditionally non-marketable risks have been made at least partly marketable during recent decades: via securization, creation of asset-backed securities, markets have been created for previously non-marketable risks such as credit risks in loan books of banks and mortgage institutions (MBSs, Mortgage Backed Securities – how beneficial the consequences of these developments have been, is, of course, open to debate), catastrophe risk of (re)insurers (catastrophe bonds), mortality and longevity risks of insurers (mortality and longevity bonds). However, these new markets are at an early stage of development, and hence less deep and liquid than well established financial markets.

References

[1] Berglund, R. Practical Financial Aspects of Insurance with Emphasis on Life Insurance. Lecture notes, Åbo Akademi University, 2003.

[2] Bühlmann, H. *Mathematical Methods of Risk Theory*. Springer, 2007.

[3] Daykin, C., Pentikäinen, T., and Pesonen, M. *Practical Risk Theory for Actuaries*. Chapman and Hall, 1994.

[4] Directive. 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II). Official Journal of the European Union 17.12.2009, 2009.

[5] Edelstein, R. H., and Quan, D. C. How Does Appraisal Smoothing Bias Real Estate Returns Measurement? Working paper, 2005.

[6] Hull, J. C. *Options, Futures and Other Derivatives, 5th Edition*. Prentice Hall, 2002.

[7] Idzorek, T. M. Developing Robust Asset Allocations. Working paper, 2006.

[8] Lehto, O. *Reaalifunktioiden teoria*. Limes r.y., 1975.

[9] Luenberger, D. G. *Investment Science*. Oxford University Press, 1997.

[10] McNeil, A., Frey, R., and Embrechts, P. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, 2005.

[11] Møller, T., and Steffensen, M. *Market-Valuation Methods in Life and Pension Insurance*. Cambridge University Press, 2007.

[12] Pesonen, M., Kauppi, M., and Mäkinen, M. *Kuolevuustarkasteluja SHV-tutkintoa varten*. artikkelikokoelma, 2005.

[13] Pesonen, M., Soininen, P., and Tuominen, T. *Henkivakuutusmatematiikka*. Yliopistopaino, 1999.

[14] Spiegel, M. R., and Liu, J. *Mathematical Handbook of Formulas and Tables, 2nd Edition*. McGraw–Hill, 1999.

[15] Stiglitz, J. E. *Freefall. America, Free Markets, and the Sinking of the World Economy*. Norton, 2010.

[16] Taleb, N. N. *Black Swan: the Impact of the Highly Improbable*. Random House, 2007.

[17] Tuomikoski, J., Sorainen, J., and Kilponen, S. *Lakisääteisen työeläkevakuutuksen vakuutusteknikkaa*. Eläketurvakeskuksen käsikirjoja 2007:4, 2007.

[18] Wütrich, M. V., Bühlmann, H., and Furrer, H. *Market-Consistent Actuarial Valuation*. EAA Lecture Notes, Springer, 2008.