OPTIMAL RETIREMENT INVESTMENT STRATEGIES

Martin Stricker

July 14, 2003

CHAPTER I

INTRODUCTION

Financial security in retirement is one of the most important financial considerations in everyone's life. Financial insecurity expresses itself as loss of income, additional expenses, insufficient income, or uncertainty of income (Rejda, 1994). It is caused by an economic loss that can be originated from death or disability of primary wage earner, illness or accident causing medical expenses, automobile accident, or perils such as fire, earthquake, wind, etc. (Carpenter, 2000).

Financial security is also a major global issue in public policy, as documented by the 1994 World Bank book "Averting the Old Age Crisis" (World Bank, 1994). This has become especially important now, as the large generation born following World War II, so called Baby Boom generation, plans to retire, and the generations following it are smaller, and thus less able to carry the burden of providing retirement income to the retirees.

While we are able to work, we can earn money to cover our cost of living; although, we always face possible unemployment, disability, illness and other personal circumstances. But how do we manage to survive when we are no longer able to work due to old age?

Before the creation of the industrial society, children were responsible to care for their parents, just as parents cared for their children during childhood.

Societal changes started with the creation of the industrial age, which followed the Industrial Revolution, started in Great Britain in the 18th century. The Industrial

Revolution then spread all over the world in the 19th century, changing the relationship between parents and children, and the role of the extended family. People started to work in factories and became separated from their families. This was also the time when societies began introducing retirement and pension systems.

Since that time, the pension systems have developed in different ways all over the globe, always achieving a common purpose: providing income in retirement. Despite differences between various countries, all pension systems in one form or another provide protection against risk of lost income after a worker stops working in a full-time position, or begins phased retirement.

If we consider retirement systems, additional questions appear automatically:

How long do we live after retirement? The increase in longevity is a major concern of pension systems. A reason for this phenomenon is probably the permanent development of our health care systems and the increase in our standards of living. Since the beginning of the 20th century, life expectancy of a newborn has increased from 47 years to approximately 76.9 years in the United States (Allen et al, 2003). Additionally, the elderly profit from the progress in medical research. One resulting effect is the absolute and relative increase in the population of persons age 65 and over. In 1900, there were roughly 3 million persons age 65 and over in the United States, whereas there were about 35 million persons of that age in 2000. The proportion of the U.S. population age 65 and over was about 4% in 1900, whereas it was about 12.7% in 2000. The age distribution of those age 65 and over has also significantly changed. A generation ago, 68% of them were 65 to 74 years old, 27% were 75 to 84 years old, and only 5% were 85 or older. Today's elderly population is shifted toward the upper end of the scale: Approximately 52.2% are 65 to 74 years old, 35.4% are 75 to 84 years old, and 12.4% are 85 or over (Allen et al, 2003).

How much money does one need to retire comfortably? To prevent outliving of one's financial resources, people should estimate what their financial needs will be during retirement. This is a difficult issue, even for a person who has just reached retirement age. The desired standard of living upon retiring serves as the goal of everybody's retirement planning. After considering how well off one wants to be upon retiring, individuals can then determine how much wealth they need to attain the desired standard of living.

The financial needs during retirement are influenced by many factors, such as support for any dependents like a disabled child, debts to be repaid like mortgage or college tuition, and lifestyle like travel, housing and hobbies. Retirement income is also affected by medical costs, which are usually at the highest point at the end of life. Therefore, retirees with health insurance that continues after retirement do not need the same amount of retirement income as those who need financial resources for their medical costs.

Once people have determined where they want to be financially during retirement, their next step is to consider how to achieve the financial goals. This process involves among other things:

- Estimation of future benefits from Social Security and pension plans, which depends on the pension system.
- Estimation of expected investment results.
- Determination of the additional amount needed to achieve financial retirement goals.
- Consideration of the effects of inflation on the cost of living.

CHAPTER II

PENSION SYSTEMS

2.1 General Pension Systems

Most developed economies in the world have gradually evolved towards a system that meets the risk of economic insecurity at an advanced age through a combination of the following:

1. social insurance programs
2. employer-sponsored pension plans (whether private or public plans)
3. personal savings
4. earnings from employment after retirement
5. income from family members and other financial security programs

The first three items are usually referred to as the *three tiers of the retirement system*, also commonly referred to as the *three-legged stool of economic security*, or the *pillars of economic security.*

All three tiers are of main importance since they complement each other. If one tier is missing in the retirement system, the remaining parts cannot compensate for it. Thus, it is important to involve all parts in a retirement system and especially in its own personal retirement plans.

1. Social insurance programs are usually government-sponsored public retirement programs that are funded by general tax revenue, payroll tax revenue, or both. If

the program is funded by payroll taxes, there is generally a relationship between the benefits and contributions. Social insurance programs may include benefits for death or disability.

2. In the employer-sponsored programs, the employer provides benefits to its employees for the time post retirement. These plans are either private or public and can be voluntary or mandatory in different countries. Tax purposes may also play an important role in these programs, depending on different regulations.

   Formal retirement income arrangements were primarily developed in the 20th century, but pensions have already existed since antiquity. Individuals who served the state, like the Roman legionnaires or the sailors of England, were pensioned after years of service.

3. Personal savings are an important third tier of the retirement system. Individuals postpone their consumption from the working period to retirement. Generally, personal savings come from any money which is left after paying taxes and after paying for its cost of living. In many countries, tax benefits are available for retirement savings. The next section discusses the effect of various tax treatments on the retirement income.

4. Depending on the personal and financial situation, retirees may keep on working in a part-time position, receiving income after retirement. A gradual process of reducing employment time from full-time, through part-time, to no employment, is called *phased retirement*. In the case of the outliving of ones financial resources, retirees sometimes have to go to work during retirement.

5. The last component of retirement financial security is income from family members and other financial security programs. In financial distress, retirement

income may also come from dependents, but this is not a preferable situation.

2.1.1 Tax Treatment of Retirement Programs

Contributions, earnings on contributions, and benefits paid are the three financial components of retirement programs which can be taxed. Each of them is either taxed or exempt from tax. The tax treatment can vary among the different distribution systems social security, private pensions, and personal savings.

- Social Security: All of the OECD (Organization for Economic Cooperation and Development) countries exempt employer distributions from tax. They tax the benefits paid (Scahill, 2002). The member countries of the OECD are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Ireland, Italy, Japan, Luxembourg, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States.

- Private Pensions: The general tax treatment among the OECD countries is the following: contributions are tax deductible or taxed at a lower rate, the earnings are tax-exempt or taxed at a lower rate, and the benefits are taxable. Only New Zealand differs from this basic approach in that they tax contributions and earnings on the retirement fund, but benefits paid are tax-exempt (Scahill, 2002).

- Retirement Savings: According to the World Bank Report (World Bank, 1994), OECD countries have granted tax advantages to retirement saving and annuity accounts set up in banks, life insurance companies, and mutual funds.

In the US, people can invest in an individual retirement account (IRA) with pre-tax money. In this case, the individual can deduct the IRA contribution before paying taxes (Scahill, 2002).

2.1.2 Types of Pension Plans

There are many different types of pension plans all over the world. To give an overview of the most important ones, we classify the types of pension plans as Gajek and Ostaszewski (upcoming in 2003).

Within the first criterion plan financing, we distinguish two different plans:

- *pay-as-you-go plans* (pension plans without a fund)
- *funded pension plans* (pension plans with a fund)

A pay-as-you-go plan is an unfunded plan, which does not accumulate assets to secure benefits of plan participants. It is financed by future income and is only backed by the credit of the plan sponsor. The reason that plan participants depend on the employer for the security of the benefits, increases their risk and is a sufficient reason to exclude pay-as-you-go plans from tax reductions in the United States. In case of bankruptcy, claims of plan participants would become worthless. In spite of having a priority in the pool of all creditors' claims by the U.S. bankruptcy law, bankruptcy proceedings are of such duration that most plan participants don't make demands on it.

Public pay-as-you-go plans are by far the most common formal system and they are the responsibility of the government, which finances public pensions (World Bank, 1994). Public pension systems do not raise solvency concerns, as the taxing power of government assures availability of funds to pay benefits.

In the United States, pay-as-you-go plans are a form of social insurance plans, such as OASDI (Social Security).

Funded pension plans avoid the possible pitfall of pay-as-you-go plans and back the benefits, which leads to an increase in the security of participants' benefits.

Qualified pension plans in the United States are defined as those that meet the requirements of ERISA and receive favorable tax treatment from Internal Revenue Code. The major tax benefits are (Hallman and Hamilton, 1994):

- contributions to the plan by the employer are tax deductible
- employees are not considered to have taxable income from the investment earnings until they receive benefit payments
- distributions from the plan can receive favorable income tax treatment

In the United States, qualified pension plans must be funded.

The next aspect in classifying pension plans is the difference between

- *defined benefit plans*, and
- *defined contribution plans*.

In a defined benefit plan, the retirement benefits are defined in advance, as the name implies. The level of the benefits is often dependent on years of employment and salary, usually a percentage of average wages of the last few years. The investment risk has the plan sponsor, who is expected to have adequate funds to provide the pension benefits. If there are fewer assets than liabilities in the plan, the premiums, in this context also called normal cost, are adjusted.

In contrast to defined benefit plans, defined contribution plans determine only the contributions made in advance. The pension benefit is dependent on the performance and the future investment return of the plan; so, the individual bears the investment risk. If the plan is privately managed, the investment risk is increased because workers have not the same competence as investment companies. In case of public management, the government may limit investment options, which introduces political risk with investment risk.

Defined benefit plans keep the option to cover disability and longevity risks, whereas it is not designated in defined contribution plans.

Another criterion for classification of pension plans is to distinguish between:

- *Occupational (employer-sponsored) plans* (pension plans without a fund), and
- *public plans* (pension plans with a fund).

Occupational (employer-sponsored) plans are privately managed and offered by employers to their employees. They are often intended to assist, attract, and retain desired workers. These programs often take advantage of tax privileges and they are mostly regulated by the government. According to the research report "Averting the Old Age Crisis" of the World Bank (World Bank, 1994), more than 40% of workers are covered by occupational schemes in the following countries: Germany, Japan, the Netherlands, Switzerland, the United Kingdom, and the United States. Worldwide, the plans tend to be defined benefit and they are at least partially funded.

In *public plans*, contributions change as necessary to provide promised benefits to the elderly. Therefore, society bears the risk of a poor economy and increased longevity. Benefit formulas usually depend on the actual economic situation. If public plans are financially in good shape and the economy is strong, benefits are often increased by the government. Otherwise, if the plan runs into deficits, pension benefits may be reduced. A recent example of this is the increase of the retirement age in Germany and the United States.

According to World bank research report (World Bank, 1994), public plans are by far the largest and most prevalent formal system. They are used in all OECD, Eastern Europe and in most Latin American countries. They are mandatory and usually redistributive defined benefit plans, that are largely funded on a pay-as-you-go basis, and are publicly managed.

Financing the pension plans can be done solely by the employer, solely by the employee, or jointly by both parties. A plan is called *contributory*, if employee financing is required. In the public sector, employees are required to make an allowance toward their defined benefit programs, which is usually made on a pretax basis. In the private sector, plans are rarely contributory. Public and private defined contribution plans are more similar in that both allow pretax contributions.

2.2 Pension System in the United States

The first industrial pension plan in the United States was introduced by the American Express Company in 1875. Shortly after, the Baltimore and Ohio Railroad Company started the second formal plan in 1880 (Allen et al, 2003). During the next half century, private pension plans grew rapidly. Most of the 400 plans were found in the railroad, banking, and public utility fields, whereas insurance companies started operating in this sector not until 1921: Metropolitan Life Insurance Company entered the market by issuing the first group insurance contract (Black, 1955). Another milestone happened in the pension business in 1974, when the Employee Retirement and Income Security Act (ERISA) was passed, which established a major foundation from an investment, tax, legal and actuarial viewpoint.

ERISA introduced the requirement that pension plans, which had advantages accrued from tax purposes (qualified pension plans), had to have a fund available for paying the benefits. Additionally, these pension plans were required to pay plan normal cost, as calculated by an actuary (Gajek and Ostaszewski, upcoming in 2003). ERISA also established the basis of the individual retirement account (IRA). The IRA was designed for individuals, which were not covered under a qualified retirement plan (Allen et al, 2003).

2.2.1 The First Tier: Social Security System

In the United States, the first tier of the three tier system of economic security is covered by the Social Security System (Old Age Security and Disability Income, or OASDI). The term Social Security refers in the United States to the cash benefits provisions of the OASDI program, whereas in international context it includes all other types of governmental programs protecting of economic risks such as unemployment, short-term sickness, work-related accidents and diseases, and medical care costs. Social Security provides the foundation of retirement protection in the United States. Currently, OASDI coverage applies for more than 90 percent of the total workforce of the United States. One half of the remaining individuals have protection through a special government retirement system. The other half are either very low paid intermittent workers or unpaid family workers (Allen et al, 2003). OASDI coverage applies also for self-employed persons. The program started originally in 1935.

Social Security benefits are financed by a payment of 6.2% of earnings up to the maximum taxable wage base by both the employer and employee. The wage base is $84,900 for 2002 and increases annually by the rate of increase in average wages in the U.S. economy (Scahill, 2002).

To be eligible for benefits as a retiree, a person must have a certain number of quarters of coverage (QCs or coverage credits), depending on the year of an individual's attainment at age 62. For fully insured status, the number of necessary QCs is computed by the formula

$$\text{number of QCs} = \text{age minus 22}$$

and has a minimum value of six QCs and a maximum of 40 QCs (Bluhm, 2000). In 1998, an employee earned one quarter of coverage for each \$700 of covered earnings in the calendar year, up to a maximum of 4 credits per year.

Retirement benefits are based on the primary insurance amount (PIA), whose computing methods are quite difficult, because several different methods are available. The first step is to calculate the average indexed monthly earnings (AIME), which is on a career-average earnings basis. The AIME is based on the average of the 35 highest years of indexed earnings. Indexing the actual earnings means increasing them to reflect nationwide wage inflation from the particular year up to the year of attaining age 60. Detailed information about computing the AIME is given in Allen et al (2003). The PIA is now computed by a three-part benefit formula, that varies for each annual cohort of persons attaining age 62.

As an example, we consider the composition of the PIA of a person turning 62 in 2002 as in (Scahill, 2002). The PIA is

- 90% of the first \$592 of AIME, plus
- 32% of the AIME in excess of \$592 through \$3,567, plus
- 15% of the AIME in excess of \$3,567.

The social security full retirement age, also called normal retirement age (NRA), was first set at age 65, which is still eligible for those attaining this age before 2003. Otherwise the NRA is gradually increasing and becomes 67 for persons born in 1960 and later.

The Social Security benefit equals the PIA for retirees at full retirement age. The benefits can begin earlier, but not before age 62. Then the benefits are reduced by a certain percentage rate: beginning at 20% at age 62 if the full retirement age is 65, increasing up to 30%, when the NRA is 67.

Retirement benefits are paid on a reduced basis to people who work while retired from Social Security. The method to control whether an individual has substantial earnings from employment is referred to as the retirement earnings test. It only applies to earned income; income from investments or pensions is not considered. The basic idea of this test is the application of an annual exempt amount, that benefits can be paid if earnings are not in excess thereof. For 2002, benefits before NRA are reduced, if earnings exceed $11,280 (Scahill, 2002). The reduction is $1 of benefit for each $2 of excess earnings.

After full retirement age, full benefits are paid regardless of the amount of wages or the fact that the person is still working. Anybody who is working after NRA and decides not to receive retirement benefits, receives an increased benefit which varies by year of attainment at age 62. The increase factor in 2002 was 7% per year for the 5 years from age 65 to 70. After age 70, no increase factor applies.

Social Security provides not only benefits for retirement, but also for disability and survivors such as widow(er)s and spouses. Spousal benefits are one-half of the PIA, beginning at full retirement age. The reduction for earlier benefit payments is 25% at age 62 for NRA of 65, and 35% for NRA of 67 respectively. Widow(er) benefits can begin as early as age 60, but they are reduced by 28.5% independent of

NRA (Scahill, 2002).

The Medicare program is closely associated with OASDI. One part, hospital insurance (HI), covers the same employed persons as OASDI and provides hospital and related benefits. Beneficiaries are insured persons who are over the age of 65 or have been disabled for at least 2.5 years (McGill et al, 1996). The other part, supplementary medical insurance, covers all persons over 65 and all disabled persons covered by hospital insurance. It provides benefits for physician and related medical services, and is financed by individual premium payments and contributions from the federal government.

2.2.2 The Second Tier: Pension Plans

The second tier of the retirement system is represented by employer-sponsored pension plans. We distinguish between defined benefit plans and defined contribution plans and whether they are privately or publicly managed.

2.2.2.1 Public Sector Pension Plans

Pension systems for government employees are of importance in the American pension market. Plans for state and local employees include almost 13 million workers (Scahill, 2002). An additional 3 million personnel are covered by the military retirement system. The most important federally run civilian systems are the Civil Service Retirement System (CSRS) and the Federal Employees Retirement System (FERS). The CSRS is the largest of the federal plans and provides retirement, survivorship, and disability benefits to career employees of the federal government. It was established in 1920 and covers employees of the federal government hired before 1984. FERS covers the federal civilian employees hired in 1984 and later (McGill et al, 1996). In the sum, state and local employee pensions controlled over $2.4 trillion in assets in 1998 (Mitchell, and Hustead, 2001).

State and local pension programs began prior to World War II. Most of the pension programs were based on a defined contribution basis. Low investment returns and high inflation during the first half of the 20th century were the cause of moving many public plans to the defined benefit form. In recent years, the defined contribution plan gained in importance in the private sector, whereas the state and local systems still concentrate on defined benefit plans. Nevertheless, the recent transition to defined contribution plans of a few systems are signs of change (Scahill, 2002).

Public pension plans are financed by employer and employee contributions, almost always on an advanced funded basis. Sometimes supplemental benefits are funded as pay-as-you-go. In addition to contributions, a few public plans also receive income from fees or earmarked levies. Generally, the investment possibilities of public plans are diversified (Scahill, 2002).

2.2.2.2 Private Retirement Systems

In the private retirement system, employees benefit from private pensions, provided by their employers. It is a voluntary system and is of great importance for the overall retirement security system, since the financial well-being of retired workers implies the loss of responsibility of supporting less fortunate members of the community.

The private pension system is heavily regulated in the United States. Most of the defined benefit and defined contribution plans are qualified plans, which are eligible to receive favorable tax treatment. Therefore, it is important to insure that tax subsidy goes to a cross-section of employees, not only to the highly-paid. The regulations consider this item in the same way the protection of employees against employer insolvency by requiring a certain level of funding of the plan (Scahill, 2002).

The legal framework also regulates the vested right to pensions of employees after no more than seven years of service with the plan sponsor. If employees quit after seven years of service, the plan sponsor is responsible for providing information about the plan, and in particular, for assuring that participants receive promised benefits from the plan.

If defined benefit plans don't have enough money to pay promised benefits, they are insured by a plan termination insurance, which is administered by the Pension Benefit Guarantee Corporation (PBGC). The insurance premium is based on the number of plan participants and on the level of funding to the plan. In return, bankrupted companies can transfer pension liabilities to the PBGC, in order that employees get a certain level of pension benefits, which is certainly less than the full benefit.

U.S. non-qualified retirement plans do not have to comply with most of the complicated regulations that apply to qualified plans. They are often used for providing benefits to executives who have already capped their benefits by various limits.

According to Scahill (2002), the participant is often taxed on the contribution when it is made, if the benefits are funded. Investment earnings are immediately taxable. Social security tax applies to non-qualified plan benefits, but not to qualified plan benefits.

2.2.2.3 Defined Benefit vs. Defined Contribution Plans

Private defined benefit plans are very common in the United States and in 1975 they nearly monopolized the pension market. As in Canada, they must be funded. Private defined contribution plans gained in importance in the late 1970s in the U.S. when high inflation and investment risk caused the will of the employers to shift the risk to plan participants.

In defined contribution plans, employers make a contribution to an individual account of an employee. The contribution is based on the employee's wage and is limited due to non-discrimination rules. Investment choices are made by the employee, but the variety of choices is established by the employer. The latter must provide four investment choices to shift the whole investment risk to the beneficiaries (Gajek and Ostaszewski, upcoming 2003):

1. a money market equivalent (very short term bonds, considered to be cash equivalent)
2. diversified portfolio of bonds
3. balanced portfolio (stocks and bonds)
4. a diversified portfolio of stocks

In the U.S., it is common to pay all benefits as lump sums in defined contribution plans.

According to Allen et al (2003), we can identify six basic types of defined-contribution plans in the United States, which are all regulated under ERISA:

- Money Purchase Pension Plans
- Profit Sharing Plans
- Savings Plans
- Employee Stock Ownership Plans (ESOPs)
- Cash or Deferred Plans under Section 401(k)

- Section 403(b) Plans

Defined benefit plans define the retirement benefits in advance rather than the contribution rate. There are four basic types of formulas which are used to establish the amount of retirement benefit each employee receives:

- The flat amount formula provides the same amount to all employees regardless of age, earnings, or length of service. It is also common to include a service requirement to recognize years of service, meaning that employees with fewer years of service receive reduced benefits.

- The flat percentage of earnings formula is designed to provide 30 to 60 percent of an employee's earnings and is used in many pension plans. The percentage might apply to *career average earnings*, i.e. employee's average earnings while participating in the plan, or to *final average earnings*, i.e. average earnings over the last few years prior to retirement. A flat percentage of earnings formula does not consider years of service.

- A flat amount for each year of service formula provides a benefit of a fixed amount per month, multiplied by the number of years of service. It is common to require a minimum number of hours that must be worked in one year to receive full credit.

- A fixed percentage of earnings for each year of service formula determines the benefits as a percentage of earnings for each year of service.

In practice, combinations or variations of the four above are used.

The majority of the benefits are retirement benefits that are paid as monthly annuities or as a lump sum. Some defined benefit plans pay disability benefits and death benefits if the participant dies before retirement. In the United States, defined benefit plans must pay a benefit to the surviving spouse of a plan participant who dies prior to retirement. Benefits are also paid to participants who leave the defined benefit plan before becoming eligible for retirement. The participant becomes vested under the plan.

Since the mid-1980s, there is a trend away from defined benefit plans toward defined contribution plans in the U.S.

There are many reasons for that change. The common criticisms of employers concerning defined benefit plans are the variability of contribution amounts and the burdensome administration of such plans, especially in time, effort, and cost.

Another reason for the seminal shift from defined benefit to defined contribution plans is the uncomplicated transfer within defined contribution plans. If an employee resigns from a job, the pension benefit will be calculated on a shortened period of service and on the last wages, in case of a defined benefit plan. This causes depreciation in benefits, especially if there are periods of high inflation between termination and retirement. In defined contribution plans, the participant's fund balance can be transferred to the new employer's plan or to an individual retirement account (IRA).

Defined contribution plans are attractive particularly to young workers and mobile workers. So, they help employers attract and retain employees, as the workforce becomes more mobile.

An important element in the decrease in defined benefit plans is that plan sponsors in the U.S. have to terminate the plan in order to withdraw surplus assets. Once a plan is terminated, employers often replace it by defined contribution plans to avoid the investment risk. To get a hold on this situation, Congress added a 20% to 50% excise tax on top of the corporate income tax (around 40% including both federal and state taxes) (Scahill, 2002).

There are also reasons to sponsor a defined benefit plan. Some companies continue to use this pension plan by habit, while other companies need them to be competitive, since in their industry, defined benefit plans are prevalent. Some employers seek to have minimum supply of their employees and use defined benefit plans as a form of retirement insurance.

The trend of the shift from defined benefit toward defined contribution plans is illustrated in the following table (EBRI, 1997).

Table 1: Shift from Defined Benefit toward Defined Contribution Plans

Primary Retirement Plan		
Year	Defined Benefit	Defined Contribution
1985	169,540	346,014
1990	112,890	498,093
1991	101,585	506,066
1992	88,400	530,627
1993	83,517	532,542

Contributory defined contribution plans are common in the United States and Canada, but contributory defined benefit plans do not exist in the United States, since employee contribution would not be tax-exempt. However, they are quite common in Canada, where they are generally tax-exempt.

2.2.2.4 Hybrid Plans

Hybrid plans are the innovative reaction to some of the legislative complications associated with defined benefit plans. They blend features of traditional defined benefit pension plans and defined contribution plans to meet the needs of employers and workforces. There are many different types and variations in the plan design of hybrid plans. But for tax purposes, each of them is considered as either a defined benefit or a defined contribution plan. Some defined benefit hybrids provide a portable lump sum benefit option to appeal to younger and more mobile workers and to keep some of the security of a defined benefit plan for older workers. Others are designed rather for old workers by using an age-weighted formula. On the other hand, defined contribution hybrids are sometimes designed to provide more assurance that income replacement goals will be met while keeping defined contribution flexibility. Other defined contribution hybrids provide a larger allocation of contributions to workers.

As a representative of defined benefit hybrid plans, we have a closer look at cash balance plans, and for defined contribution hybrid plans, we consider target benefit plans respectively.

- The cash balance plan is a defined benefit plan because it defines the benefit rather than the contribution to be made to the plan. The first plan of its kind was created by the Bank of America in 1984. The main difference to a defined benefit plan is that the actuarial value of all benefits is presented to the worker as an account balance. This amount is composed of annual pay credits and investment earnings. The annual pay credits are established by the plan actuary and reported to the Department of Labor. A published index such as the return on one-year treasury securities is generally used to determine the investment earnings. The sponsor determines the investment strategy of the underlying assets and hopes the rate credited to the accounts will be outperformed. Then the sponsor's contribution decreases, since the excess in investment earnings funds a part of the promised benefits. The portable lump sums and early accrual benefits under cash balance plans are appreciated by employees terminating at younger ages and after fewer years of service. Workers with more years of service are likely to receive more benefits from the traditional defined benefit plans (Scahill, 2002).

- Target benefit plans are officially qualified as defined contribution plans, but their funding is based on projected retirement benefits. The amount of contributions is determined actuarially to meet income replacement targets, which are expressed by a percentage of salary near the time of retirement. Nevertheless, these targets are not guaranteed once initially set, and pension benefits depend on account values reflecting the performance of the plan investments. Thus, as in the defined contribution plan, investment risk in a target benefit plan is taken by the employees. Interestingly, target benefit plans have not been accepted by the American pension market (Gajek and Ostaszewski, upcoming 2003).

2.2.3 The Third Tier: Personal Savings

The third leg consists of a variety of personal savings. They are used to supplement public and occupational retirement income. As a special feature, we also consider individual retirement arrangements (IRA), which started in the United States in the year 1975.

Personal savings is voluntary and therefore requires personal discipline. It also clarifies that retirement savings is a deferral of consumption from the present to some future time. So, individuals are faced with the decision of how much they consume now and how much they save for retirement. This is a very difficult issue. In order to help with calculations, we walk through the mechanics for a basic example, presented in Scahill (2002).

We make the following assumptions:

- you work for a company

- your expected defined benefit pension equals 1% of final 5-year average salary times years of service
- a public social security plan replaces 15% of your final year's salary in retirement
- your salary grows 4% per year
- your personal savings accumulate at 8% per year
- you want to replace 70% of your income when you retire
- you begin participating in the company defined benefit plan at the same age as you begin saving for retirement
- we are ignoring the cost of living adjustment that might be included in social security benefits and tax issues.

Table 2 shows the personal savings needed for someone who starts saving at age 30, 40 and 50 and demonstrates the shortfall of waiting to begin saving (Scahill, 2002).

We present the corresponding key formulas and calculations for the age 30:

- Final salary projects current salary to age 64, anticipating retirement at exactly age 65:

$$60,000 \cdot (1.04)^{(64-30)} = 227,659$$

- Final 5-year average salary averages the salaries projected to ages 60 through 64:

$$60,000 \cdot \frac{1}{5} \cdot [1.04^{34} + 1.04^{33} + 1.04^{32} + 1.04^{31} + 1.04^{30}] = 210,807$$

Table 2: Different Amount of Personal Savings Needed

Current age	30	40	50
Retirement age	65	65	65
Income replaced by social security	15%	15%	15%
Current salary	60,000	89,000	131,000
Final salary	227,659	228,134	226,850
Income replaced by private pension:			
Final 5-year average salary	210,807	211,247	210,058
Years of service @ retirement	35	25	15
Annual retirement benefit	73,782	52,812	31,509
Income replacement ratio	32%	23%	14%
Income replacement needed from personal savings to reach desired income replacement	23%	32%	41%
Personal savings needed:			
Annual benefit needed from personal savings	52,362	73,073	93,009
Annuity conversion factor @ age 65	8.195800735	8.195800735	8.195800735
Lump sum needed at age 65 from personal savings	429,149	598,318	762,283
Interest only accumulation factor for flat annual savings	172.3168	73.10594	27.15211
Flat annual savings needed	2,490	8,184	28,075
Interest only accumulation factor for saving a flat percent of salary each year	270.98138	104.56597	34.28064
Percent of salary that must be saved each year	2.64%	6.43%	16.97%

- Years of service @ retirement:

$$65 - 30 = 35$$

- Annual retirement benefit is 1% of final average salary × years of service at retirement:

$$(0.01) \cdot (210,807) \cdot (35) = 73,782$$

- Income replacement ratio equals annual retirement benefit divided by final salary:

$$\frac{73,782}{227,659} \approx 0.3241$$

- Annual benefit needed from personal savings is the replacement ratio needed from personal savings times the final salary:

$$(0.23) \cdot (227,659) = 52,362$$

- Conversion factor is $\ddot{a}_{65}$ using 8% interest and UP-84 unisex mortality: 8.195800735

- Lump sum needed from personal savings is the conversion factor times the annual benefit needed from personal savings:

$$(8.1958) \cdot (52,362) = 429,149$$

- Interest only accumulation factor for flat annual savings is

$$\frac{(1.08)^{\text{years to retirement}} - 1}{0.08} = 172.3168$$

- Flat annual savings is the lump sum needed divided by the accumulation factor:

$$\frac{429,149}{172.3168} = 2,490$$

- Interest only accumulation factor for saving a flat percent of salary each year is

$$\frac{(1.08)^{\text{years to retirement}} - (1.04)^{\text{years to retirement}}}{0.08 - 0.04} = 270.98138$$

- Percent of salary that must be saved each year is

$$\frac{\left(\frac{\text{Lump sum needed at retirement}}{\text{accumulation factor for \% of salary savings}}\right)}{\text{current salary}} = 2.64\%$$

2.2.3.1 Individual Retirement Accounts (IRAs)

In the United States, more than 40 million American workers were not covered by qualified retirement plans at the beginning of 1974 (Allen et al, 2003). Thereupon, Congress enabled those individuals with earned income to establish their own retirement plans on a tax-deferred basis beginning in 1975, by passing the Employee Retirement Income Security Act (ERISA). At heart, IRAs are defined contribution-type plans, established and handled by employees, without any necessary involvement on the part of their employers. Any employee or self-employed person who has earned income from personal services is eligible to establish an IRA even if he or she is already covered by a tax-qualified plan, government plan, or certain annuities. But investment income does not qualify for such a plan. An eligible individual can make tax-deductible contributions up to a fixed dollar amount set for any given year by law or 100% of earned income for the taxable year, whichever is less. The investment income earned on the contributions are currently sheltered from income tax. They are taxed when withdrawn, generally during retirement when one's income tax rate might be reduced. In the case of premature withdrawals, a tax penalty can incur.

Over time, numerous changes regarding the workings of IRAs have been made. As a result, there now are different types of IRAs with varying characteristics, features and differing tax treatments. These various IRAs include traditional deductible IRAs, traditional nondeductible IRAs, conduit IRAs, employer-sponsored IRAs, deemed IRAs added on to employer plans, and Roth IRAs. The distinct attributes of each type can be found in Allen et al (2003).

2.2.3.2 Top 10 Ways to Beat the Clock and Prepare for Retirement

To close and summarize this chapter, we want to allude to 10 steps people should take to begin their retirement planning. It is the topic of one of the brochures published by the Department of Labor's Pension and Welfare Benefits Administration (2003) to help educate citizens about the need to save for retirement. These steps are as follows:

1. Know your retirement needs. First, we need to determine the income replacement ratio we need in retirement to maintain our standard of living. It is about 70% of the pre-retirement income; 90% or more for lower earners.

2. Find out your Social Security benefits. Social Security pays the average retiree about 40% of preretirement earnings.

3. Learn about your employer's pension or profit sharing plan.

4. Contribute to a tax-sheltered savings plan.

5. Ask your employer to start a plan.

6. Put money into an Individual Retirement Account.

7. Don't touch your savings. Savings can be like a cookie jar - it is hard to keep your hand out of it!

8. Start now, set goals, and stick to them. This step encourages people to start saving early, but it also indicates that it's never too late to start.

9. Consider basic investment principles.

10. Ask questions. The last point encourages the reader to gain more information by talking with the employer, financial advisors, banks, etc..

2.3 Pension System in Germany

German Chancellor Otto von Bismarck pioneered the first pension system in the world in 1889. It started as a disability insurance for blue collar workers, and was quickly expanded to a full-fledged retirement insurance that provided disability, retirement, and survivor benefits for essentially all workers in Germany. The 'public retirement insurance' was one of the most successful pension systems in the world and was organized as a funded system. But its reserves were destroyed twice, first during hyperinflation in 1923 and second during World War II and the subsequent currency reform. After the second world war, reserves were built up again. Initially the scaled premium method was applied. The funding periods were shortened due to the high demands on the system after the war; and the official move to a pay-as-you-go system was made in 1957. Except for a liquidity reserve amounting to one month's expenditure, the system has practically no funding today.

As opposed to other countries such as the United Kingdom and the Netherlands, which originally adopted a social security system that provided a base pension, public pensions in Germany were designed to extend the standard of living that was achieved during work life to the time post retirement. Hence, public pensions are roughly proportional to labor income averaged over one's entire life and, therefore, we call the German pension system 'retirement insurance' rather than 'social security' as in the United States.

Today, the pension system in Germany consists of three main pillars:

1. the statutory public pension system for workers and employees, that includes the institutionally separate compulsory pension system for farmers and artists. Employees of the public sector are also compulsory insured. The pensions of civil servants are paid directly from the budget and are not covered by the public pension system. There are also independent pension systems run by professional associations of doctors, lawyers, and other groups of self-employed, which are usually listed in this category, since membership is compulsory.

2. the supplementary occupational pension schemes for public and private sector employees.

3. voluntary private pension provision contracts.

2.3.1 The First Tier: Statutory Pension Scheme

The statutory pension scheme (SPS) occupies a dominant position within the old-age provision system in Germany. In 2000 the SPS was made up of approximately 34.5 million actively insured individuals, which is over 80% of the labor force. Last year 23.4 million pensions were paid, of which approximately 16 million were old-age pensions (Deutsche Bank Research, 2003).

The SPS is not any less prominent in comparison of benefits relating to the different pillars of old-age provision. In Germany, over 80% of employees' retirement income comes from the first pillar. The share contributed by occupational pension schemes, the second pillar, lags far behind at around 5%. The remaining share of roughly 13% attributes to the third pillar, the private old-age provision by means of saving for retirement.

The total benefits offered by SPS in 2002 were roughly EUR 230 billion. The lion's share of the budget, 88.7%, is attributed to cash payments in the form of pension payments, which includes benefits for partial disability and surviving dependents for widows and orphans. Added to this are approximately 6% for health insurance subsidies and nursing care insurance subsidies for retirees of 0.75%. The SPS also offers related benefits in the way of medical and professional rehabilitation, which becomes 2% of the budget. Administration and procedural expenses account for 1.6% (Deutsche Bank Research, 2003).

The extensive benefits offered by SPS are financed by two sources. Two-thirds of the total expenditure is covered by individuals' contributions, which are shared equally between insured employees and their employers. These contributions are mandatory on gross wages up to the income limit for chargeable contributions, which are EUR 5,100 per month since January 1, 2003 (EUR 4,250 in East Germany). The remaining one-third are subsidies and refunds made by the Federal Government which cover expenses associated with social and redistribution policies. These expenses are called non-contribution-backed benefits, which the Federation of German Statutory Pension Institutes (VDR) defines as payments which are "not or not fully covered by insured individuals' contributions".

Since 1957 and the introduction of the SPS in its current form, the total amount of expenses is now forty times larger, having increased from just under EUR 6 billion to roughly EUR 230 billion. This development of SPS spending exceeds GDP growth substantially. In terms of GDP, SPS expenses have increased from 5.1% in 1960 to today's rate of 10.6%. The four main reasons for this are (Deutsche Bank Research, 2003):

- the politically motivated increase in pension payments. The scope of the SPS was extended both in terms of quality and quantity with the 1972 pension reform. The pension level, i.e. the net replacement rate, was increased from 63% at the beginning of the 70s to more than 70% ten years later, which has been maintained up to the present.

- the effects of population aging (Boersch-Supan, 2000). All industrialized countries are aging, and especially Germany. Germany will have the oldest population in the world in the year 2035, measured as the share of persons aged 60 and older. The proportion of German elderly will increase from 21% in 1995 to 36% in the year 2035, according to Bos et al. (1994). As a consequence, the ratio of elderly to working age persons, also called the old age dependency ratio, will increase from 21.7% in 1990 to 49.9% in 2030 in Germany. The OECD projects an increase from 20.6% in 1990 to 39.2% in 2030 for its European member countries.

- the increase in the period of time pensions are drawn. Since the introduction of the pension scheme in 1957, the length of pension payments has risen from just under ten to over 16 years, which reflects increased life expectancies. In 1957, the life expectancy of each new-born male was just under 67, and 72 for new-born females respectively. Today, life expectancies are 74.8 and 80.8.

- the German re-unification. In 1992, the pension rights of the west German federal states were transferred to the former East states. Today, approximatively 5 million of the 23.4 million state pensions are paid to recipients in East

Germany. In East Germany, only about 50% of expenses are covered by contributions, which is a consequence of low employment levels. The other half is compensated by a larger extent of the Federal Government's subsidy (EUR 7.2 billion in 2001), SPS balancing payments from West Germany (EUR 5.8 billion), and a deficit of EUR 6.5 billion.

Since the end of the 1970s, social policy has attempted to stop increased spending, which was achieved by means of spontaneous savings measures,with the exception of the pension reform in 1992. However, all savings schemes were unable to stabilize pensions financing. Social policy has been forced to repeatedly raise the contribution rate over the past few decades. The original contribution rate of 14% in the years from 1957 to 1967 has gone up to 19.5% today; it even hit 20.3% in the years 1997/98. Over the same period, the income limit for chargeable contributions was raised from roughly EUR 4,600 in 1957 to EUR 61,200 p.a..

Overall, it was the Federal Government's subsidy to the scheme that increased the greatest. State funds granted to the SPS rose from EUR 2 billion in 1957 to EUR 72.20 billion in 2002. In shares of the SPS's income, these funds surged from 25% to current levels of roughly 32%.

The 1992 pension reform was the first reaction to the demographic challenges which have been obvious for more than 20 years. A self-control mechanism was built into the pension formula with the effect that higher contribution rates reduce the pensions' adjustment factor, ensuring that retirees also contribute to the stability of pension financing.

Chancellor Kohl's government had intended to reorganize the pension system with its 1999 pension reform, which was passed at the end of 1997, focusing on the expansion to the pension adjustment formula by a demographic correction factor linked to the development of life expectancy. In this way, anticipated financial burdens caused by increasing life expectancies would be distributed equally between future retirees and contribution payers. The new government elected in autumn 1998 suspended this part of the reform and then dropped it entirely.

The red-green coalition implemented its own stabilizing concept including a new pension formula as a part of the 2001 pension reform. According to this new formula, pensions follow gross salaries, and account is taken of contributions made by members of the active population for old-age provision. If the level of contribution rises, pension increases are accordingly smaller, which is ensured by the so called gross salary adjustment factor. Until 2010, the correction factor takes account of the changes in SPS contribution rates and new increasing target rates for private old-age provision saving. The new type of adjustment should reduce the pension level to roughly 68% by 2030. This means a 10% decline would occur in pension levels to approximately 64% without reductions in the active population's net income, as a result of contributions to private old-age provision (Deutsche Bank Research, 2003).

With the 2001 reform, policy makers have abandoned the illusion that state pensions are secure and sufficient to guarantee an adequate level of income in retirement. Clear acknowledgement of additional private pension provisions superseded the state's pension assurances, and the government set the course for a more balanced old-age provision system with the entry into state-sponsored private old-age provision.

However, the benefits under the SPS, a target pension level of roughly 68% net (42% gross) and a relatively stable contribution rate until 2030, remain ambitious and are based on optimistic presumptions on decisive demographic and economic trends.

Calculations by Bonin (2001) of the Institute for the Study of Labor in Bonn show that the financing of the statutory pensions are not yet secure. The Ruerup Commission, assigned by the German government to make recommendations for social security reform, also anticipates that contribution rates will surge to over 24% by 2030.

2.3.2 The Second Tier: Occupational Pension Provision

Occupational pensions in Germany could be provided through five means, which comprise:

- direct promises covered by book reserves
- support funds (with and without reinsurance)
- direct insurance
- pension fund of insurance type
- pension funds

The last three forms are eligible for collecting the state support under the actual rules. The recent pension reform enacted in 2001 has improved the conditions for occupational pensions as vesting periods were reduced and portability was made more easy.

Social partners plan to extend coverage of occupational pensions with a preference toward pension funds. First negotiations between social partners to set up pension funds have started for a few industries. The first movers, the construction industry, the metal processing and electrical industry, and the chemical industry, have concluded agreements, but implementations of these schemes are at at very early stage.

The German pension funds are insurance companies under supervision of the insurance supervisory institution, which is still under criticism. Pension funds seem to be more similar to investment funds and should therefore be regulated under investment fund law. Many details of the pension fund regulation were rather incomplete, but in the meantime these issues were resolved through government orders. Investment rules have been defined. There is no cap on equity investing for pension funds, whereas traditional insurance companies and pension trust funds are not allowed to invest more than 30% of their portfolio into equities. A higher proportion of equities could result in higher returns and lower costs. According to Sailer (2002), regulation should follow the so called prudent person rule.

2.3.3 The Third Tier: Pension Provision Contracts

Individual access of the supplementary pension provision schemes, which were implemented in 2002, is conditional based on concluding so-called pension provision contracts (PPC). PPC can take the form of

- Private pension plans
- Insurance capitalization products
- Investment fund certificates
- Bank saving plans

and is a form of an individual pension saving account. PPCs have to meet the following requirements and conditions to be eligible for tax subsidies (Sailer, 2002):

- The plan should accumulate funds to cover the risk of old age income maintenance, and old age pension delivery must not start earlier than at age 60.
- The accumulated pension capital has to be paid out in the form of a monthly pension or monthly fixed or variable annuity with a life pension starting at age 85 to cover the longevity risk. No lump-sums can be made.

- Companies offering PPCs have to inform the client, prior to contract conclusion, concerning the fees for concluding or dissolving the contract, managing the assets and the marketing cost. They also must disclose information of the individual's account and of the volume and structure of the investment portfolio.
- Companies offering PPCs also must pledge that at least the nominal value of the contributions will be available when the contract is due to annuitisation. The right to switch to an alternative PPC with the same or a different company must be included in the contract.

PPCs could be offered by insurance companies, investment funds, banks or financial intermediaries.

2.4 Pension Systems Worldwide

This chapter is based on the World Bank book "Averting the Old Age Crisis" (World Bank, 1994).

Income insecurity in old age is a worldwide problem. In Africa and parts of Asia, the elderly present only a small part of the population and have been cared for by extended families. As urbanization, mobility, wars, and famine weaken extended family and community ties, these informal systems feel the strain. The strain is worst in those countries that are aging quickly, as a consequence of medical improvements and declining fertility. For these countries, the challenge is to move toward formal systems of income security without accelerating the decline in informal systems and without shifting more responsibility to the government than it can handle.

Latin America, Eastern Europe, and the former Soviet Union can no longer afford the formal old age security programs they introduced many years ago. Liberal early retirement provisions and generous benefits require high contribution rates, leading to widespread tax evasion. The large informal sector of the economy in many Latin American countries, for example, reflects an attempt by workers and employers to escape these high tax wages. The resulting labor market distortions reduce productivity, pushing contribution rates and then evasion even higher. The resulting limited long-term saving and capital accumulation dampen economic growth, enabling the public system to pay promised benefits.

The book "Averting the Old Age Crisis" (World Bank, 1994) gives details about the aging population:

Table 3: Percentage of Population over Sixty Years Old, 2000 - 2150

Economy	2000	2020	2050	2075	2100	2125	2150
OECD:							
simple average	20.0	26.9	31.3	30.2	30.4	30.7	30.9
weighted average	19.9	27.0	31.2	30.1	30.4	30.8	31.0
Latin America and Caribbean:							
simple average	8.8	12.0	23.7	27.9	29.4	30.2	30.6
weighted average	7.7	12.2	23.5	27.7	29.3	30.2	30.6
Eastern Europe and former Soviet Union:							
simple average	15.6	20.2	26.6	28.8	29.8	30.4	30.8
weighted average	17.0	21.5	26.5	28.7	29.8	30.4	30.8
Middle East and North Africa:							
simple average	6.8	11.1	18.3	25.4	28.8	30.0	30.5
weighted average	6.0	8.0	14.5	21.6	27.3	29.5	30.3
Sub-Saharan Africa:							
simple average	5.0	5.5	11.2	20.8	26.1	28.3	29.5
weighted average	4.4	4.9	9.9	18.8	27.7	28.1	29.4
Asia:							
simple average	7.3	11.6	20.7	25.9	28.3	29.6	30.3
weighted average	8.3	12.3	22.1	26.2	28.3	29.6	30.3

Most of these countries have cut the cost of benefits by allowing inflation to erode the real value of benefits. When Chile faced these problems in the late 1970s, it revamped the structure of its system. Other Latin American and Eastern European countries have followed them, including Mexico in 1997. Their challenge is to devise a new system and a transition path that is acceptable to those who are near retirement age, while also being sustainable and growth-enhancing to the young.

The OECD countries face similar problems as their population age and their productivity stagnates. During the past three decades of prosperity, public old age security programs covering almost the entire population paid out large pensions as poverty declined faster among the old than among the young. But over the next two decades, payroll taxes are expected to rise by several percentage points while the benefits decline. That creates intergenerational conflict between old retirees who are getting public pensions and young taxpayers who may never recoup their contributions.

Many OECD countries are moving toward a system that combines publicly managed pension plans designed to meet basic needs with privately managed occupational pension plans and/or personal savings accounts to satisfy the higher demands of middle- and upper-income groups. The challenge is to ensure the reforms are good for the whole country in the long run, even if this involves taking expected benefits away from some groups in the short run.

Pension and health care systems are strongly interrelated. In general, richer countries have better and more expensive medical facilities that increase longevity, which increase the demands on the pension system. Conversely, aging increases health care cost since health problems and medical technology are concentrated on the elderly. In Australia, public health expenses per person over age 65 is six times the spending per person under age 15, whereas the ratio is more than 10 in Hungary.

The quality of health care influences the number of people living to old age. In China, an efficient system of rural health clinics has helped raise life expectancy above that in other countries at a similar income level. Medical improvements have raised total pension costs. By raising the health level and productivity of the young, they simultaneously increase the ability of the economy to pay the higher pension costs.

In Africa and parts of Asia, old people live traditionally with their children in extended families. All family members contribute to the productivity of the household and the family group covers the consumption needs of all members. As the productive capabilities of the old decline, they are supported by the work of younger family members.

These informal systems fail for those who do not have children, those whose children have died or moved away, and those whose children do not earn enough to support less productive household members. Urbanization and increased mobility further weaken these informal systems as nuclear families replace extended families, medical progress extends life expectancy for the old, and the formalization of jobs makes it difficult for people to continue working as they age and their productivity declines.

In some African countries, informal systems have been subject to additional pressure from famines, wars, and AIDS, which have reduced the size of the working age population. East Asian informal systems are under stress because the population is aging at an unprecedented high rate. As a result of all these factors, China, Indonesia, Sri Lanka, and several African countries are considering fundamental changes in their old age security programs, but they must strike a delicate balance between moving toward formal systems of income maintenance without speeding the breakup of informal support systems, and without shifting too much responsibility to government.

The government should get involved in old age security and provide a framework for retirement income security when the informal old age security arrangements break down. Government involvement is usually justified on grounds that private capital and insurance markets are inadequate and that redistribution is needed to help the poor. Unfortunately, these interventions have introduced inefficiencies of their own and have redistributed to the rich in some cases.

About 40% of the world's workers and more than 30% of the old are covered by formal arrangements for old age, buttressed by government policy. Public spending on old age programs as a percentage of GDP (gross domestic product) increases closely with per capita income and even more closely with the share of the population that is old. If past trends continue, public spending on old age security will escalate sharply in all regions of the world over the next fifty years. The most rapid increase in public spending on old age security will occur in countries that may not expect it, because their populations are currently young.

Government intervention can take many other forms besides taxes and income transfers. For example, the government may regulate the private pension system, mandate saving, guarantee benefits, offer tax incentives, create a legal system for reliable financial institutions, and dampen inflation to encourage voluntary savings. The government can also provide information concerning the amount people need to save to provide sufficient income for retirement.

Governments can begin by encouraging families to care for their elderly. Malaysia gives a rebate to adult children whose parents live with them and an additional tax deduction for the parents' medical expenses. In Australia, France, and Switzerland, the government gives tax incentives for employers to provide pensions for their workers. Investment and insurance companies that provide annuities can be required to meet specified fiduciary standards, as in the United Kingdom and the United States.

Formal old age programs and public spending on these programs increase with economic development. Approximately 40% of the world's labor force participates in formal programs of old age income support, most of which are publicly mandated and publicly managed. Richer countries cover a larger portion of their citizens under formal plans, in part because informal systems no longer function adequately and in part because richer countries are better able to enforce compliance.

The following table specifies the manner described (World Bank, 1994), whereby GOV denotes government spending for the same year:

Table 4: Public Pension Spending Indicators (Percentages)

Country	Public pension spending / GDP	Public pension spending / GOV	Pension spending for public employees/ total public pension spending	Earmarked payroll tax minus benefit spending / benefit spending
OECD:				
average	9.2	24.7	19.3	-29.5
Latin America and Caribbean:				
average	2.0	8.5	43.6	122.9
Eastern Europe and former Soviet Union:				
average	8.0	19.1	-	-35.7
Middle East and North Africa:				
average	2.8	7.9	41.8	89.2
Sub-Saharan Africa:				
average	0.5	1.8	67.1	204.9
Asia:				
average	1.9	9.6	67.7	222.0

In high-income OECD countries, nearly all workers and their survivors are covered. In developing countries, coverage may be limited to workers in privileged occupations such as the civil service or the military. Agricultural workers, domestic servants, temporary labor, and self-employed people are the majority of the labor force in developing countries and are generally not covered. About half of the population is covered by formal pension plans in middle-income countries such as Argentina, Brazil, Costa Rica, Malaysia, and Tunisia. Coverage falls to less than 20% in lower-income countries like Indonesia, Morocco, the Philippines, and Sri Lanka, and to less than 10% in most of Sub-Saharan Africa. The world's poorest countries such as Bangladesh, Mozambique, Sierra Leone, and Somalia, have programs only for public employees, and even these are not always implemented.

Public spending on old age programs as a portion of the country's GDP is tied closely to per capita income, but it correlates even more closely with the proportion of the population that is old. In fact, spending on old age programs increases slightly faster than the proportion of the population over age 60. Austria has the highest ratio of public spending on old age programs to GDP (15%). The other OECD and Eastern European countries spend between 6% and 12%. Middle-income, middle-aged Latin American countries spend from 3% to 8% of GDP on old age programs while most young, low-income developing countries spend less than 2% on these programs.

Developing countries have less governmental capacity to collect taxes, implement complex programs, and regulate to correct these market failures than industrial countries. Similarly, the greater ease of evasion and the larger size of the informal market reduces the effectiveness of public retirement income security programs in developing countries. As a result, extended family arrangements continue to play an important role in developing countries. Since they have younger populations than industrial countries, old age security looks like a distant problem that can be ignored. It also tempts politicians to offer large pensions today to influential older workers. These reactions need to be resisted to avoid an eventual old age crisis.

Social Security programs directly concern the young as well as the old, since, depending on their design, they can either help or hinder economic growth. The programs affect the welfare of the old and the young by determining their portion of the national pie, and influencing the overall size of the national pie, respectively. Policy decisions should consider both, whether the policy will be good for the elderly and whether it will be good for the whole economy.

These are examples of how social security policy affects the entire nation:

- The payroll tax for pensions is 30% or more in most Eastern European countries, and is sometimes supplemented by general revenues. These high taxes discourage employment and deter investment in important public goods.
- In Hungary, 98% of those over age 60 are either retired or working in the informal sector to avoid payroll taxes. The remaining 2% are officially in the labor force.
- In Argentina, before the 1994 reform, 3 workers were needed to support every 2 pensioners, since early retirement increased the number of retirees and evasion decreased the number of contributors.
- In Egypt, Peru, Turkey, Venezuela, and Zambia, publicly managed pension funds lost between 12% and 37% in the 1980s.
- By 1990, the average OECD country spent 24% of its annual budget and over 8% of its GDP on old age, disability, and survivors' benefits. The average person spent more taxes on social security than on income. These numbers are expected to increase, as the populations age.

Because of theses problems, countries throughout the world evaluate and change their policies. Many OECD countries are decreasing their reliance on publicly managed plans and increasing the impact of privately managed voluntary and occupational plans. Some Latin American countries are introducing drastic structural changes which include mandatory saving plans. Eastern Europe countries need major reforms for their collapsing systems. African and Asian countries have younger populations and small formal retirement programs and must decide which way to go in the years ahead.

Table 5: Main Publicly Mandated Pension Scheme Design Features, 1991

Economy	Normal retirement age		Covered years required for full pension	Payroll tax for pensions		
	Women	Men		Worker	Employer	Combined
OECD:						
average	62.9	64.4	18.3	7.4	11.0	16.3
Latin America and Caribbean:						
average	58.7	60.8	13.9	-	-	10.5
Eastern Europe and former Soviet Union:						
average	55.3	60.3	25.0	0.0	28.3	25.5
Middle East and North Africa:						
average	57.8	60.4	13.1	4.8	8.3	10.6
Sub-Saharan Africa:						
average	56.0	56.2	13.8	3.6	5.6	9.1
Asia:						
average	55.6	56.5	10.4	5.1	7.9	13.0

CHAPTER III

MATHEMATICAL AND FINANCIAL METHODOLOGIES FOR RETIREMENT SYSTEMS MODELS

3.1 Finance

3.1.1 The Investment Environment

The *productive capacity* of an economy is defined as the goods and services that can be produced by its members. The *real assets* of an economy, defined by Bodie, Kane and Marcus (1999), are those that produce income, such as land, buildings, knowledge, and machines, in addition to the workers whose skills are necessary to use those resources. In the long run, it is the productive capacity that determines the material wealth of a society and is a function of the real assets of the economy.

In contrast to such real assets are *financial assets*, also commonly referred to as *securities* or *capital assets*, primary examples of which are stocks and bonds. Their values are derived from the claims on income produced by real assets. They are used by their buyers as tools to shift purchasing power from high-earnings periods to low-earnings periods of life. In other words, capital assets are promises of future consumption, often contingent, paid for with current consumption.

We distinguish between primitive and derivative securities. A *primitive security* derives its cash flows from real assets, whereas *derivative securities* have their cash flows derived from cash flows of a primary security, or another derivative security. Sometimes derivative securities are called derivative assets, or contingent claims since their values derive from or are contingent on the values of other assets.

3.1.2. Markets and Instruments

Financial markets are traditionally segmented into two parts: *money markets* and *capital markets*. Money market instruments are sometimes called cash equivalents and include short-term, marketable, liquid, low-risk debt securities. Capital markets include longer-term, riskier and more diverse securities. Following Bodie, Kane and Marcus (1999), we segment the capital market into longer-term fixed-income markets, equity markets, and the derivative markets for options and futures.

A detailed description of each financial instrument is found in investment textbooks, such as Bodie, Kane and Marcus (1999).

Money market instruments are Treasury Bills, Certificates of Deposit, Commercial Paper, Bankers' Acceptances, Eurodollars, Repurchase and Reverse Repurchase Agreements, and Federal Funds.

Fixed-income capital market instruments include treasury notes and treasury bonds, municipal bonds, corporate bonds, mortgages and mortgage-backed securities.

Equity securities are common and preferred stocks. The most standard derivative securities are options and futures.

The growth of futures, options, and related derivative markets has been a very significant development in financial markets (Hull, 1993).

3.1.2.1 Options

An *option* gives its holder the right to buy or sell an underlying for a predetermined price, called the *exercise* or *strike price*, on or before a specified expiration date. A *call option* is the right to buy and a *put option* is the right to sell an underlying. When the market price exceeds the exercise price, the holder of a call option can get the payoff equal to the difference between the stock price and the exercise price by exercising its option. If the option remains unexercised until the expiration date, the option expires and no longer has value. Therefore, call options provide greater profit when stock prices increase. On the other hand, profits on put options increase when the asset value falls. If the purchase or sale in the option contract can only happen at its expiration (or maturity) then the option is called *European*, whereas an option that can be exercised at any time before maturity is called *American*. A European call is an option to buy the underlying at its expiration date at a predetermined price, whereas a European put gives its holder the right to sell the underlying for the strike price on the expiration date.

3.1.2.2 Futures and Forward Contracts

A *forward contract* on a commodity is a contract to purchase or sell a specific amount of the commodity at a specific predetermined price and at a predetermined time in the future. Forwards are private transactions. The trader who takes the *long position* commits to purchasing the commodity on the delivery date. The *short position* is held by the trader who commits to delivering the commodity at contract maturity.

The *forward price* is an agreed-upon price which applies at the point of time of delivery. It is negotiated in that the initial payment is zero. Therefore the value of the contract is zero when it is initiated.

If we enter into a forward contract through an organized exchange, and have to put down a payment, called a margin, and every day a market price adjustment is made to our margin account, with maintenance margin required, we have entered into a *futures contract*.

3.1.2.3 Trading Securities

In the *primary market*, new securities are marketed to the public by investment bankers. Once they are issued, they are traded among private investors in the *secondary markets*, which consist of national and local securities exchanges, over-the-counter (OTC) market, and direct trading between two parties.

In the United States, the New York Stock Exchange (NYSE) and the American Stock Exchange (AMEX) are national in scope. The NYSE is the oldest and largest stock exchange. According to Bodie, Kane and Marcus (1999), the NYSE accounts for over 85% of the trading that takes place on stock exchanges. AMEX is the second largest stock exchange in the United States and focuses on listing small to mid-size corporations. It accounted for less than 5% of the trading volume during the year 1996.

Regional exchanges primarily list firms located in a particular geographic area and include Boston, Cincinnati, Midwest (Chicago), Pacific, and Philadelphia.

The over-the-counter market is for securities that are not listed on an exchange. There are no membership requirements for trading, nor are there listing requirements for the roughly 35,000 securities. OTC stock prices are listed daily in newspapers. The OTC market is the main market for bonds. The National Association Of Securities Dealers Automatic Quotations, or NASDAQ, is a computerized information network that provides brokers and dealers with price quotations on securities traded over-the-counter since 1971. NASDAQ quotes are published in the financial pages of most newspapers. According to Bodie, Kane and Marcus (1999), securities of nearly 5,500 firms are quoted on the system, which is now called the Nasdaq stock market.

Most transactions in exchange or OTC markets are the result of *market orders*. In a market order, investors buy or sell at prevailing price. A *limit order* reduces the price uncertainty of a transaction. It specifies a price at which an investor wants to buy or sell the security. A *stop order* is a delayed market order. The seller specifies a price below the current market price, at which the order becomes a market order.

A *short sale* allows investors to profit from declining security prices. An investor who wants to sell a stock, but does not own it, can borrow the stock from a broker and sell it in the marketplace. Later, the short seller must purchase a share of the same stock in the market and return it. In the meantime, any stock dividends that are paid must be credited to the lender's account.

3.1.3 Fundamental Terms of Investment Analysis

The *comparison principle* is used to simplify decision making in financial markets. We evaluate the investment by comparing it with other investments available. The financial market provides a basis for comparison.

In economics *arbitrage* means a simultaneous purchase and sale of the same item, with the purchase price below the sale price.

In finance *arbitrage* means creating an investment portfolio which does not require any capital outlay and produces no losses under any circumstances in the future, and results in income with positive probability.

A financial market is considered *efficient* if prices reflect information available about securities. We distinguish among three versions of market efficiency. A financial market is considered:

- *weakly efficient*, if prices reflect all historical information, first and foremost all historical prices, which implies that trend analysis is fruitless.
- *semi-strongly efficient*, if prices reflect all historical and current information publicly available, which implies that fundamental analysis does not work.
- *strongly efficient*, if prices reflect all information in existence, which implies that

insider trading does not work.

An important feature of financial markets is that they are *dynamic*, meaning the same or similar financial instruments are traded on a continuous basis. In this sense, the future price of an asset is regarded as a process moving in time and subject to uncertainty. There are a few standard models that are used to represent price processes. These include binomial lattice models, difference equation models, and differential equation models.

Typically, information such as records of the past prices are used to specify the parameters of such a model.

Investment is itself dynamic since markets are dynamic as well. The value of an investment changes with time, and the composition of portfolios may change.

3.1.4 Utility Functions

An investor seeks to maximize his benefit when he or she makes an investment decision. It would be easy to rank the choices, if the outcomes from all alternatives were certain. In general, the choice is not obvious. To evaluate the worth to the investor of different wealth levels, we use *utility functions*, which are defined as follows:

Definition 1 *(Panjer, 1998) A function* $u\colon \mathbb{R}_+ \to \mathbb{R}$*, with existing second derivative* u'' *is called a utility function, if the following properties hold:*

(a) $u(w)$ *is the utility of an investor at wealth* $w \in \mathbb{R}_+$*,*

(b) $u'(w) > 0$ *for all* $w \in \mathbb{R}_+$*,*

(c) $u''(w) \leq 0$ *for all* $w \in \mathbb{R}_+$*.*

For a given utility function $u(x)$, any function of the form

$$v(x) = au(x) + b, \quad \textit{with } a > 0$$

is said to be equivalent to $u(x)$.

The given definition of equivalent utility functions $u(x)$ and $v(x)$ is reasonable, since they produce the same ranking of wealth, even though the numbers are different.

Definition 2 *(Luenberger, 1998) A function f defined on an interval $[a, b]$ of real numbers is said to be concave if for any α with $0 \leq \alpha \leq 1$ and any x and y in $[a, b]$ there holds*

$$f\big(\alpha x + (1-\alpha)y\big) \geq \alpha f(x) + (1-\alpha)f(y).$$

A utility function u is said to represent risk-averse decision maker on $[a, b]$ if it is concave on $[a, b]$. If u is concave everywhere, the decision maker with that utility function is said to be risk-averse.

The fact that investors usually use concave utility functions is also referred to as the *risk aversion principle.* The degree of risk aversion is related to the magnitude of the convexity in the utility function: the greater the convexity, the greater the risk aversion. Measures of risk aversion are

- the absolute risk aversion function, also called the Arrow-Pratt absolute risk aversion coefficient, which is $R_A(x) = -\frac{u''(x)}{u'(x)}$, and
- the relative risk aversion function $R_R(x) = -\frac{xu''(x)}{u'(x)}$.

Remark 1

1. *Property* (b) *of the definition of utility functions has the meaning that the utility function is an increasing continuous function, and an investor prefers more wealth to less wealth.*

2. *Property* (c) *means that* $u(x)$ *is a concave function, and that the investor is risk averse.*

A *risk-neutral* utility function has the form $u(x) = ax + b$, with $a > 0, b \in \mathbb{R}$. An investor using this utility function ranks wealth levels only by their expected values, and no account for risk is made.

In order to rank the wealth levels of two outcome random wealth variables x and y, we compare their expected utility values $E\big[u(x)\big]$ and $E\big[u(y)\big]$. An investor is assumed to always prefer the larger value.

The measure expected utility is meaningless except for comparing alternatives. Therefore we introduce another measure: the certainty equivalent.

Definition 3 *(Luenberger, 1989) The certainty equivalent* C *of a random wealth variable* ω *is defined to be the amount of a certain, risk-free wealth that has a utility level equal to the expected utility of* ω*. In mathematical notation, the certainty equivalent* C *satisfies*

$$u(C) = E\big[u(x)\big].$$

A *risk-lover* uses a utility function that is not concave. The investor adjusts the expected return upward to get a certainty equivalent that exceeds the alternative of the risk-free investment.

The decision which utility functions should be used, depends on the investor's risk tolerance and financial environment. The most commonly used utility functions are described by Panjer (1998):

(i) Quadratic Utility:

$u(x) = x - \frac{x^2}{2b}$, for $x < b$,

$u'(x) = 1 - \frac{x}{b}$,

$u''(x) = -\frac{1}{b}$,

$R_A(x) = \frac{1}{b-x}$.

(ii) Exponential Utility:

$u(x) = 1 - e^{-ax}$, for $x > 0$ and $a > 0$,

$u'(x) = ae^{-ax}$,

$u''(x) = -a^2e^{-ax}$,

$R_A(x) = a$.

(iii) Power Utility:

$u(x) = \frac{1}{\alpha}(x^\alpha - 1)$, for $x > 0$ and $\alpha \in (0, 1)$,

$u'(x) = x^{\alpha-1}$,

$u''(x) = (\alpha - 1)x^{\alpha-2}$,

$R_A(x) = \frac{1-\alpha}{x}$.

(iv) Logarithmic Utility:

$u(x) = a\ lnx + b$, for $x > 0$ and $a > 0$,

$u'(x) = \frac{a}{x}$,

$u''(x) = -\frac{a}{x^2}$,

$R_A(x) = \frac{1}{x}$.

An investor with initial wealth ω compares two random gains G_1, G_2 by using a utility function: he or she prefers G_1 to G_2, if

$$E\big[u(\omega + G_1)\big] > E\big[u(\omega + G_2)\big] \ ,$$

that is, if the expected utility from G_1 exceeds the expected utility from G_2.

Theorem 1 *(Jensen's Inequality) Let X be a random variable whose expected value exists, and let $u''(w) < 0$ for all $w \in \mathbb{R}$. Then,*

$$u\big(E[X]\big) \geq E\big[u(X)\big].$$

Jensen's Inequality says that for any random variable G,

$$u\big(\omega + E[G]\big) > E\big[u(\omega + G)\big] \ .$$

This indicates that an investor who can choose between a random gain G and a fixed amount equal to its expectation, always prefers the latter.

3.1.4.1 Mean-Variance Analysis

The *mean-variance analysis* is one way to formalize the risk aversion principle. In this approach, the uncertainty of the return on an asset is characterized by the mean value of the return and the variance of the return. Following the risk aversion principle, a risk-averse investor will select the investment vehicle that has the smallest variance, if all investment opportunities have the same mean. This mean-variance method of formalizing risk is the basis for most of the methods of quantitative portfolio analysis, which was pioneered by Harry Markowitz (1952, 1959).

3.1.5 Typical Investment Problems

Each investment problem has unique features, but many fit into a few broad categories. We briefly outline some important problem types here.

First we consider the general *pricing problem*: given an investment with known payoff characteristics, which can be random, what is the reasonable price? In other words, how much is this investment worth today? The pricing problem is usually solved by use of the comparison principle. However, the application of that principle is not simple in most instances. We will elaborate on the pricing problem for options and assets later.

Hedging is the process of reducing the financial risks associated with normal business operations or investments. Investors take account of the interplay between asset returns when evaluating the risk of a portfolio. For that purpose, at least two investment vehicles are necessary which have offsetting patterns of returns.

There are many ways that hedging can be carried out: through insurance contracts, futures contracts, options, and special arrangements. Indeed, the major use of these financial instruments is for hedging, and not for speculation.

3.1.6 Introduction to Interest Rates

The amount of money a borrower of capital pays to a lender of capital for its use may be defined as interest. We analyze various quantitative measures of interest, as in Kellison (1991).

The *principal* is the initial amount of money (capital) invested and the total amount received after a period of time is called the *accumulated value*. The difference between the accumulated value and the principal is the *amount of interest*, or just *interest*, earned during the period of investment.

The time from the date of investment is usually measured in years and is denoted by t. For the investment of one unit of principal, the *accumulation function* $a(t)$ gives the accumulated value at time $t \geq 0$.

There are three main properties of accumulation functions:

1. $a(0) = 1$

2. $a(t)$ is generally an increasing function. Although a decrease and therefore negative interest is possible mathematically, it is not relevant to most situations in practice.

3. If interest accrues continuously, the function will be continuous.

The *effective rate of interest* i is defined as the amount of money that one unit invested at the beginning of a period will earn during the period. Interest is paid at the end of the period. In terms of the accumulation function, we have

$$i = a(1) - a(0) \quad \text{or} \quad a(1) = 1 + i.$$

Simple interest has the property that the interest is not reinvested to earn additional interest. Therefore the amount of interest earned during each period is constant. In this case the linear accumulation function is $a(t) = 1 + it$ for $t \geq 0$.

For *compound interest*, the interest earned is automatically reinvested. The associated accumulation function is $a(t) = (1 + i)^t$ for $t \geq 0$.

We are also interested in measuring the intensity with which interest is operating at each moment of time. This measure of interest over an infinitesimally small interval of time is called the *force of interest*.

For defining the force of interest, we use an amount function $A(t)$ which gives the accumulated value at time $t \geq 0$ of an original investment of k. Thus, we get

$$A(t) = k \cdot a(t) \quad \text{and} \quad A(0) = k.$$

Definition 4 *Let $A(t)$ be a differentiable amount function. Then the force of interest δ_t is defined as $\frac{A'(t)}{A(t)}$.*

As an interpretation, we arrange the equation $A(t) \cdot \delta_t = A'(t) \approx \frac{A'(t+\Delta t) - A(t)}{A(\Delta t)}$ as follows:

$$A(t) \cdot \big(\delta(t)\Delta t\big) \approx A'(t + \Delta t) - A(t),$$

that is the product of the amount invested at time t and the effective rate per period of Δt years is approximately the interest earned on $A(t)$ over the next Δt years.

In case of compound interest the force of interest

$$\delta_t = \frac{\big((1+i)^t\big)'}{(1+i)^t} = \frac{(1+i)^t \; ln(1+i)}{(1+i)^t} = ln(1+i)$$

is constant. Therefore we denote δ_t as δ or r.

In case of simple interest,

$$\delta_t = \frac{(1+ti)'}{(1+ti)} = \frac{i}{1+ti}.$$

An alternative expression for δ_t is

$$\delta_t = \frac{a'(t)}{a(t)} = \frac{d}{dt} \; ln\big(A(t)\big) = \frac{d}{dt} \; ln\big(a(t)\big).$$

We replace t by r and integrate both sides between the limits 0 and t. Thus,

$$\int_0^t \delta_r dr = \int_0^t \frac{d}{dr}\ ln\big(A(r)\big)dr = ln\big(A(r)\big)\big|_0^t = ln\left(\frac{A(t)}{A(0)}\right).$$

Hence, we get

$$e^{\int_0^t \delta_r dr} = \frac{A(t)}{A(0)} = \frac{a(t)}{a(0)} = a(t).$$

In practice, the force of interest is often a constant, whereby in theory, it may vary instantaneously. Let the force of interest be constant over n periods, and let n be a positive integer. Then we get

$$\begin{aligned} e^{\int_0^n \delta_t dt} &= e^{n\delta} \quad \text{since } \delta_t = \delta \text{ for } 0 \leq t \leq n \\ &= a(n) = (1+i)^n \end{aligned}$$

so that

$$e^{\delta} = 1 + i$$

or

$$i = e^{\delta} - 1.$$

The latter expresses the functional relation between δ and i. Equivalently,

$$\delta = ln(1 + i).$$

3.2 Pricing of Capital Assets

The *rate of return* is defined as

$$\text{rate of return} = \frac{\text{amount received - amount invested}}{\text{amount invested}}.$$

The prospective rate of return of an asset is usually uncertain. Accordingly, we consider the rate of return to be a random variable, which is described by its expected value μ and by its standard deviation σ. Investors prefer more of μ over less and less of σ over more.

A *risk-free asset* has a return that is deterministic; it is known with certainty and therefore has $\sigma = 0$.

3.2.1 Introduction to Portfolio Theory

3.2.1.1 Portfolios of one Risky and one Risk-free Asset

We begin by considering a portfolio of one risky asset and one risk-free asset in general to introduce basic terms and principles of portfolio theory.

Since we apply the procedures to the case of more assets, we denote the portfolio of one risky asset by P and the risk-free asset by F. We assume the composition of the optimal risky portfolio P is known.

The basic concern is what portion γ of the investment budget should be allocated to the risky asset P. The remaining portion $1-\gamma$ is invested in the risk-free asset F. We denote the rate of return on the risky asset by r_P, its expected rate of return by $E(r_P)$ and its standard deviation by σ_P. The rate of return on the risk-free asset is denoted by r_f.

The *risk premium* on the risky asset is defined as $E(r_P) - r_f$, the difference of the expected rate of return on the risky and the risk-free asset.

We denote the combination of the two investment vehicles as the complete portfolio C. Its rate of return r_C is

$$r_C = \gamma r_P + (1-\gamma) r_f$$

Taking the expected value, we get

$$\begin{aligned} E(r_C) &= \gamma E(r_P) + (1-\gamma) r_f \\ &= r_f + \gamma[E(r_P) - r_f] \end{aligned}$$

Thus, the expected rate of return on the complete portfolio consist of two parts: The rate of return on the risk-free asset, also considered as the base rate of return for any portfolio, and additionally a risk premium that depends on the risk premium of the risky asset and the portion of the budget invested in it. Since investors are considered to be risk averse, they will not invest in a risky portfolio with a negative risk premium.

The complete portfolio standard deviation equals the risky standard deviation multiplied by the weight of the risky asset in that portfolio:

$$\sigma_C = \gamma \sigma_P$$

We can also plot the portfolio characteristics as a function of γ in the expected return-standard deviation plane. The vertical axis describes the expected return on the complete portfolio $E(r_C)$ and the horizontal axis describes the standard deviation σ_C. Since the standard deviation of the risk-free asset is zero, F is the y-intercept. The risky asset P is plotted with its standard deviation and expected return. If an investor invests solely in the risky asset, then $\gamma = 1$ and the resulting portfolio is P. For $\gamma = 0$, the resulting portfolio is the risk-free portfolio F.

The midrange portfolios are lying on a straight line connecting points F and P, whereby γ lies between zero and 1. To derive the equation, we substitute $\gamma = \sigma_C/\sigma_P$

in the above equation:

$$\begin{aligned} E(r_C) &= r_f + \gamma[E(r_P) - r_f] \\ &= r_f + \frac{\sigma_C}{\sigma_P}[E(r_P) - r_f] \end{aligned}$$

Thus the expected return of the portfolio is a function of its standard deviation with intercept r_f and slope $S = \frac{E(r_P)-r_f}{\sigma_P}$. This straight line is called the *capital allocation line (CAL)*, which depicts all risk-return combinations available to investors. Its slope is the increase in the expected return of the portfolio per unit of additional standard deviation.

If investors can borrow at the risk-free rate r_f, they can construct portfolios that may be plotted on the CAL to the right of P in the expected return-standard deviation plane. However, nongovernment investors cannot borrow at the risk-free rate. The risk of a borrower's default causes higher interest rates. In practice, borrowing to invest in the risky portfolio is enforced by having a margin account with a broker.

3.2.1.2 Portfolios of two risky assets and a risk-free asset

Now, we extend the scenario to the case of two risky assets. We are faced with the portfolio construction problem for the two risky securities, and later we will combine it with the risk-free asset.

First we determine the risk-return opportunities available from the set of risky assets. They are described by the *minimum-variance frontier* of risky assets which is a graph in the expected return-standard deviation plane. This frontier gives the lowest possible portfolio variance for a given portfolio expected return. For any targeted expected return, we can calculate the minimum-variance portfolio, using the set of data for expected returns, variances, and covariances.

Since all individual assets lie to the right of the frontier, portfolios of only one asset are inefficient. The portfolio with the smallest standard deviation is called *global minimum variance portfolio* and corresponds to the leftmost point on the line. The part of the frontier that lies above this point is called the *efficient frontier* and marks portfolios which are candidates for the optimal portfolio. They have the highest value of μ for a given σ, and the lowest value of σ for a given μ respectively.

The next step is to combine the portfolios on the efficient frontier with risk-free asset and to decide which combination is optimal. Therefore we consider two portfolios A and B on the efficient frontier. As described earlier, we get the capital allocation line for each portfolio. They have the same intercept r_f, but different slopes:

$$S_A = \frac{E(r_A) - r_f}{\sigma_A} \quad \text{and} \quad S_B = \frac{E(r_B) - r_f}{\sigma_B} \ .$$

Since the slope is the measure of extra return per extra risk, the optimal portfolio is the one with the steepest slope of the CAL. The resulting CAL is tangent to the efficient frontier.

The last step is to determine the appropriate mix between the optimal risky portfolio P and the risk-free asset. Basically, we are interested in the portion γ of the investment budget which is allocated to the risky portfolio P. It depends on the nature of the investor's risk aversion. Investors who are risk averse reject investment portfolios that have a zero or a negative risk premium. They "penalize" the expected rate of return of a risky portfolio by a certain amount or percentage to account for the risk involved. The greater the risk, the larger is the penalty. In contrast, risk-neutral investors rank the investments only by their expected rate of returns. The level of risk is irrelevant and there is no penalty for it.

The formalization of this risk-penalty system is based on a *utility* score, that each investor assigns to the investment portfolio based on its expected return and risk. Portfolios are described by higher utility scores for higher expected returns and by lower scores for higher volatility. The following utility value U is commonly used by financial theorists for a portfolio with expected return $E(r)$ and standard deviation σ (Bodie, Kane and Marcus, 1999):

$$U = E(r) - (0.005)A\sigma^2 \ ,$$

where A is the index of the investor's risk aversion. Investment portfolios are seen as equally attractive if they have the same utility score. An increase in standard deviation lowers utility and therefore has to be adjusted by an increase in expected return. These portfolios lie on a curve in the expected return-standard deviation plane, namely the *indifference curve*.

If we consider the complete portfolio, the investor wants to maximize his or her utility level U, by choosing the best proportion for the risky asset, γ. Mathematically, we have the following problem:

$$\begin{aligned} \max_{\gamma}(U) &= \max_{\gamma}(E(r_C) - (0.005)A{\sigma_C}^2) \\ &= \max_{\gamma}(r_f + \gamma[E(r_P) - r_f] - (0.005)A\gamma^2{\sigma_C}^2) \end{aligned}$$

The first derivative with respect to γ is

$$E(r_P) - r_f - (0.01)\gamma A{\sigma_P}^2 \ .$$

Setting this expression equal to zero and solving for γ yields to the optimal position for risk-averse investors in the risky asset, γ^*:

$$\gamma^* = \frac{E(r_P) - r_f}{(0.01)A\sigma_P^2}$$

Thus, the optimal position in the risky asset is inversely proportional to the level of risk and risk aversion, and directly proportional to the risk premium of the risky asset.

3.2.2 Capital Asset Pricing Model

One basic approach to the pricing of securities are Equilibrium Pricing Models: Prices are such that the overall economy is in equilibrium. The capital asset pricing model is a representative of equilibrium pricing models. It was developed primarily by Sharpe (1964), Lintner (1965) and Mossin (1966) on the basis of the Markowitz mean-variance portfolio model. The following approach is presented in Bodie, Kane and Marcus (1999).

We consider an equilibrium model with the following assumptions:

1. There are many investors, each with small wealth compared to the total wealth of all investors. Individual investors are price-takers, i.e. security prices are unaffected by an investor's trade.

2. All investors have the same time horizon. This is myopic in that it ignores everything that might happen after the end of the single-period horizon.

3. All investments are publicly traded financial assets. Investors may borrow or lend any amount at a fixed, risk-free rate.

4. Investors pay no taxes on returns and no transaction costs on trades in securities. In reality, we know that investors are in different tax brackets and this may influence their investment decisions. Furthermore, commissions and fees

are costly and depend on the size of the trade and the good standing of the individual investor.

5. Information is costless and available to all investors.

6. All investors are rational mean-variance optimizers. They prefer more returns, and less standard deviation of returns, and seek to maximize μ and minimize σ.

7. All investors have homogeneous expectations. They analyze securities in the same way and share the same economic view of the world. This results in identical estimates of the probability distribution of future cash flows from investing in the available securities.

These assumptions ignore many real-world complexities. However, we can gain some powerful insights into the nature of equilibrium in security markets.

To describe the first conclusion of this model, we first introduce the *market portfolio (M)*. It includes all traded assets in the economy and equals the sum over the portfolios of all individual investors. For simplicity, we refer to all risky assets as *stocks.* The CAPM implies that all investors will hold the same portfolio of risky assets in proportions that duplicate representation of the assets in the market portfolio.

This is clear, since if all investors use identical Markowitz analysis (assumption 6) applied to the same assortment of securities (assumption 3) for the same time horizon (assumption 2) and use identical estimates of the probability distribution of future cash flows (assumption 7), they all must arrive at the same optimal risky portfolio. The CAPM guarantees that all stocks are included in the optimal portfolio: If all investors avoid one stock, the demand is zero and the price takes a free fall. As the stock gets cheaper, it becomes attractive enough to be included in the optimal portfolio.

Second, the market portfolio will be the tangency portfolio to the optimal capital allocation line derived by each of the investors. The result is called the *capital market line (CML)*, which is the best attainable capital allocation line from the risk-free rate through the market portfolio M. All investors hold M as their optimal portfolio, differing only in the amount invested in the risky portfolio and in the risk-free asset.

We use the following notation:

- M is the market portfolio
- i represents an individual security
- r_i is the rate of return on security i
- $E(r_i)$ is the expected rate of return on i
- r_f denotes the rate of return on the risk-free asset
- γ is the portion invested in the risky portfolio, and $1-\gamma$ is the portion invested in the risk-free asset
- ${\sigma_M}^2$ is the variance of the market portfolio
- A^* is the coefficient of risk aversion, $A^* = 0$ for risk-neutral investors; higher levels of risk aversion are expressed in larger values for A^*
- $\bar{A}^*$ is the average degree of risk aversion across investors

The *risk premium* on the market portfolio is defined as the difference of the expected rate of return on the market portfolio and the rate of return on the risk-free asset, $E(r_M) - r_f$. It depends on the average risk aversion of all market participants.

Since any borrowing position must be offset by the lending position of the creditor in the CAPM economy, the net borrowing and lending across all investors must be zero. Consequently the average position in the risky portfolio is 100%, or $\bar{\gamma} = 1$. As mentioned earlier, each investor chooses a proportion γ for investing in the optimal portfolio M, such that

$$\gamma = \frac{E(r_M) - r_f}{(0.01)A^* \, {\sigma_M}^2} \; .$$

Setting $\gamma = 1$ in this equation, we get the implication of the CAPM that the risk premium on the market portfolio is proportional to its risk and the average degree of risk aversion of the investors:

$$E(r_M) - r_f = \bar{A}^* \, {\sigma_M}^2 (0.01)$$

The risk premium on an individual asset $E(r_i) - r_f$ is a function of its covariance with the market, and is proportional to the risk premium on the market portfolio M, and the *beta coefficient* of the security. Beta is a normalized version of the covariance of the asset with the market portfolio and measures the extent to which returns on the stock and the market move together. Beta is defined as

$$\beta_i = \frac{Cov(r_i, r_M)}{{\sigma_M}^2}$$

and the risk premium on an individual asset is

$$\begin{aligned} E(r_i) - r_f &= \frac{Cov(r_i, r_M)}{{\sigma_M}^2} \, [E(r_M) - r_f] \\ &= \beta_i \, [E(r_M) - r_f] \end{aligned}$$

Following Luenberger (1998), we state the Capital Asset Pricing Model as a theorem.

Theorem 2 *Under the assumptions stated above, the expected return $E(r_i)$ of any asset i satisfies*

$$E(r_i) - r_f = \beta_i \left[E(r_M) - r_f\right] ,$$

if the market portfolio is efficient.

Proof:

Consider a portfolio consisting of the market portfolio M and an asset i. For any α, let α be the portion invested in i, and $1-\alpha$ be the portion invested in M. $\alpha < 0$ is possible and considered as borrowing at the risk-free rate. The rate of return of the portfolio, r_α is

$$r_\alpha = \alpha\ r_i + (1-\alpha)\ r_M\ .$$

The expected value and the standard deviation are as follows:

$$\begin{aligned}
E(r_\alpha) &= \alpha E(r_i) + (1-\alpha)E(r_M) \\
\sigma_\alpha &= \sqrt{Var(\alpha r_i + (1-\alpha)r_M)} \\
&= \sqrt{\alpha^2\sigma_i^2 + (1-\alpha)^2\sigma_M^2 + 2Cov(\alpha r_i, (1-\alpha)r_M)} \\
&= \sqrt{\alpha^2\sigma_i^2 + 2\alpha(1-\alpha)Cov(r_i, r_M) + (1-\alpha)^2\sigma_M^2}
\end{aligned}$$

As α varies, these values describe a curve in the expected return-standard deviation plane, which is tangent to the capital market line CML at the point M corresponding to the market portfolio. The curve cannot cross the CML. We use this tangency condition to derive the formula.

We have

$$
\begin{aligned}
\frac{dE(r_\alpha)}{d\alpha} &= E(r_i) - E(r_m) \\
\frac{d\sigma_\alpha}{d\alpha} &= \frac{\alpha\sigma_i^2 + (1-2\alpha)Cov(r_i, r_M) + (\alpha - 1)\sigma_M^2}{\sigma_\alpha}\ .
\end{aligned}
$$

Thus,

$$
\left.\frac{d\sigma_\alpha}{d\alpha}\right|_{\alpha=0} = \frac{Cov(r_i, r_M) - \sigma_M^2}{\sigma_M}\ .
$$

Using the relation

$$
\frac{dE(r_\alpha)}{d\sigma_\alpha} = \frac{dE(r_\alpha)/d\alpha}{d\sigma_\alpha/d\alpha}
$$

we obtain the slope of the portfolio curve at M:

$$
\left.\frac{dE(r_\alpha)}{d\sigma_\alpha}\right|_{\alpha=0} = \frac{\big(E(r_i) - E(r_M)\big)\sigma_M}{Cov(r_i, r_M) - \sigma_M^2}
$$

This must equal the slope of the CML. Hence,

$$
\frac{\big(E(r_i) - E(r_M)\big)\sigma_M}{Cov(r_i, r_M) - \sigma_M^2} = \frac{E(r_M) - r_f}{\sigma_M}\ .
$$

Solving for $E(r_i)$ leads to the stated formula:

$$
E(r_i) = r_f + \left(\frac{E(r_M) - r_f}{\sigma_M^2}\right) Cov(r_i, r_M) = r_f + \beta_i\big(E(r_M) - r_f\big)
$$

□

3.2.2.1 Conclusions

In terms of the risk premium, the CAPM states that the risk premium of an asset is proportional to the risk premium of the market portfolio. The proportion factor is

β.

Another interpretation comes to the conclusion that the risk premium of an asset is directly proportional to its covariance with the market. Furthermore, it is the covariance that determines the risk premium of the asset.

From the CAPM view it follows that the risk of an individual asset is fully characterized by its beta. A greater beta implies greater expected return. Betas of most U.S. stocks range between 0.5 and 2.5 according to Luenberger (1998). The beta of the market portfolio equals 1 by definition.

An asset that is uncorrelated with the market and therefore has a $\beta = 0$, has an expected rate of return equal to the risk-free rate r_f. More extreme is an asset with a negative value of β. In this case, $E(r_i) < r_f$ leads to an asset with expected rate of return less than the risk-free rate, even though the asset may have very high risk, which is measured by its σ. Such an asset reduces the portfolio risk when it is combined with the market and therefore provides a form of insurance. It does well, when everything else does poorly.

3.3 Option Valuation

An investor can buy or sell options. If the security price is likely to rise, an investor should purchase a call or sell a put. Conversely, if the security price is likely to fall, then a put should be purchased or a call should be sold. The relation between the value of a put and a call with the same exercise price on a given stock is given by the put-call parity.

3.3.1 Put-Call Parity

We make the following assumptions:

1. There are no dividends on the underlying.

2. We can borrow and invest at the risk-free rate r (force of interest).

3. There are no transaction costs, no taxes.

4. Short selling and borrowing is unrestricted. Securities can be traded in fractional amounts.

5. No arbitrage is possible.

We consider a European call and a European put on a stock. The notation is as follows:

- Time now = t
- Time at expiration = T
- Price of the stock at the current time t = S_t
- Exercise price = K
- Value of European put = P
- Value of European call = C
- Payoff of the European call at expiration is
$$C_T = \left\{ \begin{array}{cc} S_T - K & \text{if } S_T \geq K \\ 0 & \text{if } S_T < K \end{array} \right\} = max(S_T - K, 0) = (S_T - K)^+$$
- Payoff of the European put at expiration is
$$P_T = \left\{ \begin{array}{cc} 0 & \text{if } S_T > K \\ K - S_T & \text{if } S_T \leq K \end{array} \right\} = (K - S_T)^+$$

The portfolio consisting of a long call and a short put costs $C - P$ at time t. It has a payoff of

$$\begin{aligned} C_T - P_T &= (S_T - K)^+ - (K - S_T)^+ \\ &= \left\{ \begin{array}{rl} S_T - K & \text{if } S_T \geq K \\ 0 & \text{if } S_T < K \end{array} \right\} - \left\{ \begin{array}{rl} 0 & \text{if } S_T > K \\ K - S_T & \text{if } S_T \leq K \end{array} \right\} \\ &= S_T - K \end{aligned}$$

at time T. We can also obtain the payoff $S_T - K$ at time T without using options: we buy the asset, which costs S at time t, and borrow $Ke^{-r(T-t)}$. The net cost is $S - Ke^{-r(T-t)}$ and the resulting payoffs at time T are identical to the portfolio considered. Since there are no arbitrage trades, we have

$$C - P = S - Ke^{-r(T-t)},$$

which is the *put-call parity* relation for European options, often written as

$$C + Ke^{-r(T-t)} = S + P.$$

3.3.2 Black-Scholes Option Pricing Model

We consider a European call option on a non-dividend-paying stock, with exercise price K, and the option is exercised $\tau = T - t$ years from now. The current price of the underlying (typically a stock) is $S_t = S$. The stock price on the exercise date is $S_T = Se^{\tau R}$, where R is the instantaneous rate of return. We assume that the random variable R is normally distributed with variance $\sigma^2 \tau$, and all assumptions for the put-call parity apply.

Then, the *Black-Scholes formula* for the value of a European call is given by

$$C = S \cdot N(d_1) - Ke^{-r\tau} \cdot N(d_2),$$

where

$$\begin{aligned}
N(x) &= \text{standard cumulative normal probability distribution} \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy \\
r &= \text{risk-free force of interest, continuously compounded} \\
d_1 &= \frac{ln(S/K) + r\tau + \frac{1}{2}\sigma^2\tau}{\sigma\sqrt{\tau}} \\
d_2 &= d_1 - \sigma\sqrt{\tau}
\end{aligned}$$

Since $ln(\frac{1}{Ke^{-r\tau}}) = -ln(Ke^{-r\tau}) = -ln(K) - ln(e^{-r\tau}) = ln(\frac{1}{K}) + r\tau$, we get

$$\begin{aligned}
d_1 &= \frac{ln\left(\frac{S}{Ke^{-r\tau}}\right) + \frac{1}{2}\sigma^2\tau}{\sigma\sqrt{\tau}}, \text{ and} \\
d_2 &= \frac{ln\left(\frac{S}{Ke^{-r\tau}}\right) + \frac{1}{2}\sigma^2\tau}{\sigma\sqrt{\tau}} - \sigma\sqrt{\tau} \\
&= \frac{ln\left(\frac{S}{Ke^{-r\tau}}\right) + \frac{1}{2}\sigma^2\tau}{\sigma\sqrt{\tau}} - \frac{\sigma^2\tau}{\sigma\sqrt{\tau}} \\
&= \frac{ln\left(\frac{S}{Ke^{-r\tau}}\right) - \frac{1}{2}\sigma^2\tau}{\sigma\sqrt{\tau}}.
\end{aligned}$$

From the put-call parity we get

$$\begin{aligned}
P &= C - S + Ke^{-r\tau} \\
&= S \cdot N(d_1) - Ke^{-r\tau} \cdot N(d_2) - S + Ke^{-r\tau} \\
&= S \cdot (N(d_1) - 1) - Ke^{-r\tau} \cdot (N(d_2) - 1)
\end{aligned}$$

By using the property $N(d_1) - 1 = -N(-d_1)$ of the standard normal distribution, we get the formula for the value of a European put:

$$P = Ke^{-r\tau} \cdot N(-d_2) - S \cdot N(-d_1)$$

3.3.2.1 Derivation of the Black-Scholes Formula

In a *risk-neutral world*, the price of a security is the expected present values of its cash flows. Let $g(S_T)$ be the probability density function of the stock price at a future time T in a risk-neutral world. As in Hull (1993) we assume that S_T has a lognormal distribution:

$$ln(S_T) \sim \Phi\left[ln(S) + \left(r - \frac{\sigma^2}{2}\right)\tau, \sigma\sqrt{\tau}\right],$$

where $\Phi(m, s)$ denotes a normal distribution with mean m and standard deviation s.

The value for a European call is the expected present value of its cash flows, i.e. the excess of the asset price at expiry over the exercise price, but not less than zero:

$$\begin{aligned} C &= E\left[e^{-r\tau} max(S_T - K, 0)\right] \\ &= e^{-r\tau} \int_K^{\infty} (S_T - K)\, g\,(S_T)\, dS_T \end{aligned}$$

Since S_T has a lognormal distribution and the probability density function (p.d.f.) of a random variable with mean μ and standard deviation σ is

$$\frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(ln(x)-\mu)^2},$$

the p.d.f. of S_T is as follows:

$$g(x) = \frac{1}{x\sqrt{2\pi\sigma^2\tau}} e^{-\frac{1}{2\sigma^2\tau}\left(ln(x)-ln(S)-\left(r-\frac{\sigma^2}{2}\right)\tau\right)^2}$$

Therefore we get for the value of a European call:

$$\begin{aligned} C &= e^{-r\tau}\int_K^\infty (x-K)g(x)dx \\ &= e^{-r\tau}\int_K^\infty \frac{x-K}{x\sigma\sqrt{2\pi\tau}} e^{-\frac{1}{2\sigma^2\tau}\left(ln(x/S)-\left(r-\frac{\sigma^2}{2}\right)\tau\right)^2} dx \end{aligned}$$

We evaluate the integral following Kellison (1991) by making a change of variable. Let

$$y = \frac{ln(x/S) - \tau(r-\sigma^2/2)}{\sigma\sqrt{\tau}}$$

so that

$$dy = \frac{1}{x\sigma\sqrt{\tau}}\, dx$$

and

$$x = Se^{\left[\tau(r-\sigma^2/2)+y\sigma\sqrt{\tau}\right]}$$

Let the lower limit of the integral after the change of variable be F, so that

$$K = Se^{\left[\tau(r-\sigma^2/2)+F\sigma\sqrt{\tau}\right]}$$

or

$$F = \frac{ln(K/S) - \tau(r-\sigma^2/2)}{\sigma\sqrt{\tau}}.$$

Thus, after the change of variable we get

$$\begin{aligned}
C &= e^{-r\tau}\int_F^\infty \frac{Se^{\left[\tau(r-\sigma^2/2)+y\sigma\sqrt{\tau}\right]}-K}{\sqrt{2\pi}}\ e^{-y^2/2}\ dy \\
&= S\int_F^\infty \frac{e^{\left[-r\tau+r\tau-\tau\sigma^2/2+y\sigma\sqrt{\tau}-y^2/2\right]}}{\sqrt{2\pi}}\ dy - Ke^{-r\tau}\int_F^\infty \frac{e^{-y^2/2}}{\sqrt{2\pi}}\ dy \\
&= S\int_F^\infty \frac{e^{\left[-\frac{1}{2}(\tau\sigma^2-2y\sigma\sqrt{\tau}+y^2)\right]}}{\sqrt{2\pi}}\ dy - Ke^{-r\tau}\int_F^\infty \frac{e^{-y^2/2}}{\sqrt{2\pi}}\ dy \\
&= S\int_F^\infty \frac{e^{\left[-\frac{1}{2}(y-\sigma\sqrt{\tau})^2\right]}}{\sqrt{2\pi}}\ dy - Ke^{-r\tau}\int_F^\infty \frac{e^{-y^2/2}}{\sqrt{2\pi}}\ dy
\end{aligned}$$

The resulting integrals represent areas in the standard normal distribution. The first integral is the area to the right of $F-\sigma\sqrt{\tau}$, or to the left of $-F+\sigma\sqrt{\tau}$ of the bell-shaped curve. Likewise, the second integral is the area to the right of F or to the left of $-F$. Thus,

$$C = S\cdot N(-F+\sigma\sqrt{\tau}) - Ke^{-r\tau}\cdot N(-F).$$

Now let

$$\begin{aligned}
d_1 = -F+\sigma\sqrt{\tau} &= \frac{ln(S/K)+\tau(r-\sigma^2/2)+\sigma^2\tau}{\sigma\sqrt{\tau}} \\
&= \frac{ln(S/K)+\tau(r+\sigma^2/2)}{\sigma\sqrt{\tau}}
\end{aligned}$$

and

$$d_2 = -F = \frac{ln(S/K)+\tau(r-\sigma^2/2)}{\sigma\sqrt{\tau}}.$$

This yields to

$$C = S\cdot N(d_1) - Ke^{-r\tau}\cdot N(d_2)$$

which is the Black-Scholes formula.

Empirical tests of the Black-Scholes formula show that the formula works reasonably well under certain conditions. According to Kellison (1991), the Black-Scholes formula tends to develop significant errors in the following situations:

1. The exercise price K is far from the current market price S.

2. For securities with volatility much above or below average (i.e. σ being quite high or low).

3. If the expiry date is far in the future.

3.4 Probability Theory and Stochastic Processes

First we introduce the mathematical framework, basic tools and terms used in this paper.

We state the following definitions, as presented by Dothan (1990):

Definition 5 *A probability space is a triple* (Ω, Σ, Pr) *where* Ω *is a sample space,* Σ *is a* σ*-field of subsets of* Ω*, and* Pr *is a probability measure on* Σ*:* $Pr\colon \Sigma \to [0, 1]$*, with*

1. $Pr(\emptyset) = 0,\ Pr(\Omega) = 1$

2. $Pr(\bigcup_{i\geq 1} A_i) = \sum_{i\geq 1} Pr(A_i)$ *for any countable collection* $\{A_1, A_2, ...\}$ *of disjoint elements of* Σ*.*

Definition 6 *The Borel* σ*-field* $\mathcal{B}$ *of* $\mathbb{R}$ *is the set of all the countable unions of open intervals and the complements of countable unions of open intervals. Any set in* $\mathcal{B}$ *is called a Borel set.*

Definition 7 *A function*

$$X\colon \Sigma \to \mathbb{R}$$

is Borel measurable on a σ-field Σ if and only if for every Borel set B the set $X^{-1}(B)$ belongs to Σ.

Definition 8 *A Borel measurable function*

$$X\colon \Omega \to \mathbb{R}$$

is called a random variable X on a probability space (Ω, Σ, Pr).

Definition 9 *A stochastic process on the probability space (Ω, Σ, Pr) is a function $X\colon [0,T] \times \Omega \to \mathbb{R}$ such that for every $0 \le t \le T$, $X(t,\cdot)$ is a random variable on (Ω, Σ, Pr). For a fixed $\omega \in \Omega$, the function $X(\cdot,\omega)\colon [0,T] \to \mathbb{R}$ is called a sample path of the process X. The realization of the stochastic process X at time t and state ω, $X(t,\omega)$, is also denoted $X_t(\omega)$. The stochastic process X is also denoted $\{X_t\}$.*

For the definition of the Wiener process, we need the property of independent increments of stochastic processes (Panjer, 1998):

Definition 10 *A stochastic process $\{X_t\}$ is said to have independent increments if all collections of random variables of the form*
$X_{t_2} - X_{t_1}, ..., X_{t_n} - X_{t_{n-1}}$ are independent, for any set of points $0 \le t_1 < ... < t_n$.

Brownian motion process (Ross, 2000) is defined as follows:

Definition 11 *A stochastic process $\{X(t),\ t \ge 0\}$ is said to be a Brownian motion process if*

(i) $X(0) = 0$;

(ii) $\{X(t),\ t \geq 0\}$ *has stationary and independent increments;*

(iii) for every $t > 0$*,* $X(t)$ *is normally distributed with mean 0 and variance* $\sigma^2 t$*.*

The Brownian motion process is sometimes called the *Wiener process.* When $\sigma = 1$, the process is called *standard Brownian motion.* We can convert any Brownian motion to the standard process by letting $B(t) = X(t)/\sigma$.

Definition 12 *We say that* $\{X(t),\ t \geq 0\}$ *is a Brownian motion process with drift* μ *and variance parameter* σ^2 *if*

(i) $X(0) = 0$*;*

(ii) $\{X(t),\ t \geq 0\}$ *has stationary and independent increments;*

(iii) $X(t)$ *is normally distributed with mean* μt *and variance* $t\sigma^2$*.*

Equivalently, let $\{B(t),\ t \geq 0\}$ be a standard Brownian motion and then define

$$X(t) = \sigma B(t) + \mu t$$

to get a Brownian motion process with drift μ and variance parameter σ^2.

Definition 13 *Let* $\{Y(t),\ t \geq 0\}$ *be a Brownian motion process with drift coefficient* μ *and variance parameter* σ^2*. The process* $\{X(t),\ t \geq 0\}$ *defined by*

$$X(t) = e^{Y(t)}$$

is called a geometric Brownian motion.

CHAPTER IV

SECURITIES MARKET MODEL WITH PRICE DENSITY

The next model we concentrate on is based on the idea that securities are generally priced by a price density. We present continuous models for the one-period securities market and the continuous time complete market, as developed by Gerber and Shiu (2000).

4.1 Basic Model-Framework

We consider an investor, who invests money for a fixed time period. We assume he starts to invest at time 0 and he wants to have an optimized fortune at a fixed future time T, whereby $T > 0$. Hence, T represents the retirement age of the investor, as we are interested in investing for retirement.

We make the assumption that there is no addition or withdrawal of funds in the time period $[0, T]$, which is unrealistic, because we usually don't have the entire amount of money for retirement available before time T. Investing for retirement is a process of saving and therefore we have additions of funds. Withdrawals are also possible, for example if we want to invest the funds in another way. But the assumption is still useful, since we can separate the time period in shorter periods, so that we have no changes in the funds.

The money is invested in the securities market. To determine the prices for the securities, we first introduce the mathematical environment. We consider a probability space (Ω, Σ, Pr), where

- $\Omega = \mathbb{R}_+$
- $\Sigma = \sigma$-Algebra on Ω, and
- Pr is the probability measure that applies to the real-world probabilities.

In general, the price for a security equals the expected present value of the payoffs. In the context of a risk-neutral world, we consider risk-free rates. We use real-world probabilities, which are assumed to be known. Additionally we use risk-adjusted interest rates that are higher than risk-free rates, since inflation and other influences are taken into account.

In the market considered, the securities or assets are priced by a price density, which is defined as follows.

Definition 14 *The price density, denoted by Ψ, gives the time-T price for a random payment Y due at the strategic planning horizon T by the expectation $E\big[Y\Psi\big]$. Ψ is a positive random variable over the probability space (Ω, Σ, Pr) with expectation $E\big[\Psi\big] = 1$.*

Remark 2 *We can also consider $E\big[Y\Psi\big]$ the forward price for the random payment Y due at time T.*

As pointed out in Gerber and Parfumi (1998), $E[Y\Psi]$ can be rewritten by

$$E\big[Y\Psi\big] = E\big[Y\big] + E\big[Y\Psi\big] - E\big[\Psi\big]E\big[Y\big] \ ,$$

and therefore the price of Y is

$$E\big[Y\big] + Cov(Y, \Psi) \ .$$

The price of the payment is its expectation modified by an adjustment expressing the market conditions.

We assume that the risk-free force of interest r remains constant over time. This is unrealistic, since interest rates are always effected by the economy and therefore are subject to variations. On the other hand, the 30 year T-bond is an investment vehicle with guaranteed interest rate for the long period of 30 years. Thus, for the simplified one period model, the assumption is still useful.

Consequentially, we can now specify the price of the random payment Y which is due at time T. The price of Y at time 0 is $e^{-rT}E\big[Y\Psi\big]$.

4.1.1 A Single Risky Asset

First we consider a securities market model consisting of only one risky asset, a risk-free asset or bond, and their derivative securities. We extend the results later to a model with multiple risky assets.

We assume that the value of the risk-free asset accumulates at the risk-free rate r and the risky asset pays no dividend in the considered time period $[0, T]$. Let $S(t)$ be the price of a unit of the risky asset at time t, $(0 \leq t \leq T)$, with all dividends reinvested. Then

$$X(t) = ln\big[S(t)/S(0)\big] \tag{1}$$

denotes the continuously compounded total rate of return over the time interval $[0, t]$. An equivalently expression is

$$S(t) = S(0)\ e^{X(t)}.$$

We assume that $X(t)$ is a Wiener process (or Brownian motion) with "real-world" drift parameter μ and diffusion coefficient σ, as in the classic Black-Scholes model. That means for each fixed t, the random variable $X(t)$ is normally distributed with mean μt and variance $\sigma^2 t$.

Gerber and Shiu (1994, 1996) show that there is a unique number h^*, such that

$$\Psi = \frac{S(T)^{h^*}}{E\big[S(T)^{h^*}\big]} \tag{2}$$

is the price density of the securities market. The number h^* is called the *risk-neutral Esscher parameter*. If we consider $S(T)$ as a random payment, it follows directly from the definition of the price density, that

$$\begin{aligned} S(0) &= e^{-rT} E\big[S(T)\Psi\big] \qquad (3) \\ &= e^{-rT} E\left[S(T)\frac{S(T)^{h^*}}{E\big[S(T)^{h^*}\big]}\right] \end{aligned}$$

If we substitute $S(0) = S(T)e^{-X(T)}$ from the defining relation of the rate of return, we get

$$\begin{aligned} 1 &= \frac{1}{S(0)}\ e^{-rT} E\left[\frac{S(T)^{h^*+1}}{E[S(T)^{h^*}]}\right] \\ &= \frac{1}{S(0)}\ e^{-rT} E\left[\frac{S(0)^{h^*+1}\ e^{X(T)(h^*+1)}}{E[S(0)^{h^*} e^{X(T)h^*}]}\right] \\ &= e^{-rT}\ \frac{1}{S(0)}\ \frac{S(0)^{1+h^*}}{S(0)^{h^*}}\ \frac{E\left[e^{X(T)(h^*+1)}\right]}{E\left[e^{X(T)h^*}\right]} \\ &= e^{-rT}\ \frac{E\left[e^{X(T)(h^*+1)}\right]}{E\left[e^{X(T)h^*}\right]} \qquad (4) \end{aligned}$$

The moment generating formula for a random variable $X \sim N(\mu, \sigma^2)$, that is normally distributed with mean μ and variance σ^2 is

$$\begin{aligned} M(h) &= E\big[e^{hX}\big] \\ &= \int_{-\infty}^{\infty} e^{hx}\frac{1}{\sigma\sqrt{2\pi}}\ exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}\ dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\ exp\left\{-\frac{1}{2\sigma^2}[x^2-2(\mu+\sigma^2h)x+\mu^2]\right\}\ dx\ . \end{aligned}$$

We use

$$x^2-2(\mu+\sigma^2h)x+\mu^2 = \big[x-(\mu+\sigma^2h)\big]^2-2\mu\sigma^2h-\sigma^4h^2\ ,$$

to get

$$\begin{aligned} M(h) &= exp\left(\frac{2\mu\sigma^2h+\sigma^4h^2}{2\sigma^2}\right)\int_{-\infty}^{\infty}\underbrace{\frac{1}{\sigma\sqrt{2\pi}}\ exp\left\{-\frac{1}{2\sigma^2}\big[x-(\mu+\sigma^2h)\big]^2\right\}}_{=\text{ p.d.f. of } X\sim N(\mu+\sigma^2t,\sigma^2)}\ dx \\ &= exp\left(\frac{2\mu\sigma^2h+\sigma^4h^2}{2\sigma^2}\right)\cdot 1 \\ &= exp\left(\mu h+\frac{\sigma^2h^2}{2}\right) \qquad \text{for } X\sim N(\mu,\sigma^2)\ . \end{aligned}$$

Thus, for $X(t)\sim N(\mu t,\sigma^2 t)$, the moment-generating formula is

$$M(h) = E\left(e^{hX(t)}\right) = exp\left(h\mu t+\frac{1}{2}h^2\sigma^2 t\right)\ . \tag{5}$$

Then we get from (5),

$$\frac{E\left(e^{(1+h^*)X(T)}\right)}{E\left(e^{h^*X(T)}\right)} = \frac{exp\left((1+h^*)\mu T+\frac{1}{2}(1+h^*)^2\sigma^2T\right)}{exp\left(h^*\mu T+\frac{1}{2}{h^*}^2\sigma^2T\right)} = exp\left(\mu T+\frac{1}{2}(1+2h^*)\sigma^2T\right)$$

which has to be e^{rT} by (4). Therefore, with

$$e^{rT} = exp\left(\mu T + \frac{1}{2}(1 + 2h^*)\sigma^2 T\right)$$

we obtain

$$\begin{aligned} h^* &= \left(rT - \mu T - \frac{1}{2}\sigma^2 T\right)\frac{1}{\sigma^2 T} \\ &= \frac{r - \sigma^2/2 - \mu}{\sigma^2}. \end{aligned} \tag{6}$$

Remark 3 *We can express the price of a payment by its expectation with respect to a modified probability measure, Q, that is defined by the relation*

$$E_Q\big[Y\big] = E\big[\Psi Y\big] \qquad \textit{for all } Y.$$

In this context, Ψ is the Radon-Nikodym derivative of the Q-measure with respect to the original probability measure (Gerber and Parfumi, 1998). With respect to Q, which is also called risk-neutral measure or equivalent martingale measure, the process $\{X(t)\}$ is still a Wiener process with changed drift parameter

$$\mu^* = r - \frac{\sigma^2}{2} \tag{7}$$

but with the same diffusion coefficient σ. Hence, (6) becomes

$$h^* = \frac{\mu^* - \mu}{\sigma^2}\ . \tag{8}$$

The numerator of $-h^*$ is the risk premium on the risky asset, $\mu - \mu^*$, which is the excess of the expected instantaneous rate of return of the risky asset over the risk-free *rate. As in Hull (1993), the quantity $-h^*\sigma$ is sometimes called the market price of risk.*

Another way to describe the price dynamics of the risky asset is in terms of a stochastic equation. In our notation we have:

$$\frac{dS(t)}{S(t)} = \left(\mu + \frac{\sigma^2}{2}\right) dt + \sigma dZ(t), \tag{9}$$

where $\{Z(t)\}$ is a standard Wiener process. Hull (1993) uses μ in place of our $(\mu + \sigma^2/2)$ term. With his notation, we would have

$$\begin{aligned} \mu &= r, \quad \text{and} \\ h^* &= \frac{r-\mu}{\sigma^2}. \end{aligned}$$

We define $Z^*(t) = Z(t) - \sigma h^* t$. $\{Z^*(t)\}$ is a standard Wiener process with respect to the risk-neutral measure. (9) changes to

$$\frac{dS(t)}{S(t)} = rdt + \sigma dZ^*(t). \tag{10}$$

Equation (10) shows that under the risk-neutral measure the expected instantaneous rate of return of the risky asset is the risk-free rate r.

4.1.2 Multiple Risky Assets

A more realistic securities market model is one consisting of a risk-free asset accumulating at rate r, n risky assets, and their derivative securities. As in the previous model, we assume that no dividends are paid in the considered time period $[0, T]$. For $k = 1, 2, ..., n$, let $S_k(t)$ denote the price of the k-th risky asset at time t,

$0 \leq t \leq T$, and

$$X_k(t) = ln\big[S_k(t)/S_k(0)\big]$$

denotes the continuously compounded total rate of return over the time interval $[0, t]$. We assume that $\{X_1(t), X_2(t), ..., X_n(t)\}$ is an n-dimensional Wiener process with the properties:

$$E\big[X_k(1)\big] = \mu_k$$

$$Cov\big[X_i(1), X_j(1)\big] = \sigma_{ij} \; .$$

Gerber and Shiu (1994) use the assumption that the covariance matrix $C = (\sigma_{ij})$ is nonsingular to show that there are n numbers $h_1^*, h_2^*, ..., h_n^*$ such that

$$\Psi = \frac{S_1(T)^{h_1^*} S_2(T)^{h_2^*} \cdots S_n(T)^{h_n^*}}{E\left[S_1(T)^{h_1^*} S_2(T)^{h_2^*} \cdots S_n(T)^{h_n^*}\right]}$$

is the price density of the securities market. If we apply the moment generating function formula for multivariate normal random variables to the generalization of (3),

$$S_k(0) = e^{-rT} E\big[S_k(T)\Psi\big], \quad k = 1, 2, ..., n$$

we can determine the n risk-neutral Esscher parameters $h_1^*, h_2^*, ..., h_n^*$. Similar to (7), we define

$$\mu_k^* = r - \frac{\sigma_{kk}^2}{2}, \quad k = 1, 2, ..., n.$$

Then we get the risk-neutral Esscher parameters by solving the system of linear equations:

$$\sum_{j=1}^{n} \sigma_{ij} h_j^* = \mu_i^* - \mu_i, \quad i = 1, 2, ..., n.$$

Remark 4 *With the row vectors*

$h^* = (h_1^*, h_2^*, ..., h_n^*)$,

$\mu = (\mu_1, \mu_2, ..., \mu_n)$, *and*

$\mu^* = (\mu_1^*, \mu_2^*, ..., \mu_n^*)$

we have a generalization of (8):

$$h^* = (\mu^* - \mu)C^{-1}.$$

4.2 Characterizing the Optimal Terminal Wealth

An investor invests in the securities market at time 0. Starting with wealth ω at this time, he wants to optimize his wealth at a certain time $T(T > 0)$. The wealth at time T of an investor after buying a random payment Y in the securities market is

$$W(Y) = \omega e^{rT} + Y - E\big[Y\Psi\big]. \tag{11}$$

Since T is the decision horizon, we call the random variable $W(Y)$ terminal wealth. The price of $W(Y)$ is

$$E\big[W(Y)\Psi\big] = \omega e^{rT} \text{ for each } Y. \tag{12}$$

We maximize the expectation $E\big[u(W(Y))\big]$ to determine the optimal terminal wealth $W(Y)$ at time T, which we denominate W_T. Directly from (12), we get

$$E\big[W_T\Psi\big] = \omega e^{rT}. \tag{13}$$

Theorem 3 *The optimal terminal wealth can be characterized by*

$$u'(W_T) = E\big[u'(W_T)\big]\,\Psi, \tag{14}$$

which is an equivalent condition for optimality.

Proof:

"$\Leftarrow$": First, we show that (14) is a necessary condition for optimality. Let W_T be the optimal terminal wealth. Furthermore, let V be an arbitrary random payment at time T and ξ an arbitrary real number. Then

$$W_T + \xi V - \xi E\big[V\Psi\big]$$

is another terminal wealth. We define the function $\Phi : \mathbb{R} \to \mathbb{R}$ as

$$\Phi(\xi) = E\big[u(W_T + \xi V - \xi E[V\Psi])\big],$$

with first derivative $\Phi'(\xi) = E\big[u'\big(W_T + \xi V - \xi E[V\Psi]\big)\ \big(V - E[V\Psi]\big)\big]$. It follows from the assumption that W_T is optimal, that

$$E[u(W_T)] \geq \Phi(\xi) \qquad \text{for all } \xi.$$

Therefore, the function $\Phi(\xi)$ has a maximum at $\xi = 0$. The first-order condition

$$\Phi'(0) = 0$$

is

$$E\big[u'(W_T)\ (V - E[V\Psi])\big] = 0$$

$$\Leftrightarrow E\big[Vu'(W_T)\big] - E\big[u'(W_T)E[V\Psi]\big] = 0$$

$$\Leftrightarrow E\big[Vu'(W_T)\big] - E\big[V\Psi\big] \cdot E\big[u'(W_T)\big] = 0$$

$$\Leftrightarrow E\big[Vu'(W_T)\big] - E\big[E[u'(W_T)]V\Psi\big] = 0$$

$$\Leftrightarrow E\big[V\big\{u'(W_T) - E[u'(W_T)]\Psi\big\}\big] = 0$$

Since V is arbitrary, the expression within the braces in the last equation has to be zero, which is equivalent to (14).

"$\Rightarrow$": Next, we show that (14) is a sufficient condition. Let $u'(W_T) = E\big[u'(W_T)\big]\Psi$. Since

$$\begin{aligned} u'(x) &= \lim_{h\to 0} \frac{u(h+x) - u(x)}{h} \quad \text{by definition of the derivative} \\ &= \lim_{y\to x} \frac{u(y) - u(x)}{y - x} \quad \text{by setting } y = x + h \end{aligned}$$

we have for the strictly concave utility function

$$\begin{aligned} u'(x) &> \frac{u(y) - u(x)}{y - x} \quad \text{for } x < y \\ \Leftrightarrow \quad & u(y) < u(x) + u'(x)(y - x) \quad \text{for } x < y. \end{aligned} \tag{15}$$

Suppose W_T satisfies the equation (14). Any other terminal wealth can be written in the form

$$W_T + Y - E\big[Y\Psi\big],$$

where Y is a non-degenerate random variable. Then

$$u\big(W_T + Y - E[Y\Psi]\big)$$

$$\leq u(W_T) + u'(W_T)\big(Y - E[Y\Psi]\big) \quad \text{by (15)}$$

$$= u(W_T) + E[u'(W_T)]\Psi\big(Y - E[Y\Psi]\big) \quad \text{by (14)},$$

with strict inequality holding with positive probability. Then, after taking expectations, we get

$$\begin{aligned} E\big[u(W_T + Y - E[Y\Psi])\big] &< E\big[u(W_T)\big] + E\big[u'(W_T)\big] \cdot E\big[\Psi Y - E[Y\Psi]\big] \\ &= E[u(W_T)] + 0, \end{aligned}$$

showing that W_T is optimal. As an additional result, we get that W_T is unique. □

The proved condition (14) states that the random variable for the marginal utility of optimal terminal wealth, $u'(W_T)$, must be a multiple of the price density Ψ. With the notation

$$m = m(\omega) = E\big[u'(W_T)\big], \tag{16}$$

(14) becomes

$$u'(W_T) = m\Psi. \tag{17}$$

We denote the inverse of the marginal utility function u' as v. v exists, since $u'' < 0$. If we apply

$$v\big(u'(x)\big) = x \tag{18}$$

to both sides of (17), we obtain

$$W_T = v(m\Psi), \tag{19}$$

which changes (13) to

$$E\big[\Psi v(m\Psi)\big] = \omega e^{rT}. \tag{20}$$

In the next section we present some important examples of the optimal terminal wealth variable W_T. In each of these examples, the formula of the functional relationship between m and ω, (20), leads to an explicit expression for m, which can be substituted in (19).

4.3 LRT Utility Functions

To facilitate understanding of optimal decision, we introduce the risk tolerance function as Panjer (1998):

Definition 15 *In the domain of the twice-differentiable utility function $u(x)$, the risk tolerance function is defined as*

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

Remark 5 *Since we assume that u is a risk-averse utility function with $u' > 0$ and $u'' < 0$, $\tau(x)$ is strictly positive.*

If two utility functions have the same risk tolerance function, we say they are *equivalent*. A utility function is a member of the class of *linear risk tolerance* (LRT) utility functions, if the corresponding risk tolerance function τ is a linear function or a constant. The LRT class is also called *hyperbolic absolute risk aversion* (HARA) class of utility functions, since the reciprocal of the risk tolerance function is the (Arrow-Pratt) absolute risk aversion function.

Following Gerber and Shiu (2000), we classify the LRT utility functions into three subclasses, depending on whether $\tau(x)$ is constant, linear and decreasing, or linear and increasing.

4.3.1 Constant Risk Tolerance Utility Functions

Constant risk tolerance utility functions are exponential utility functions with parameter $a > 0$:

$$\begin{aligned} u(x) &= -\frac{e^{-ax}}{a}, \quad -\infty < x < \infty, \\ \Rightarrow \quad & u'(x) = e^{-ax}, \quad -\infty < x < \infty, \\ \Leftrightarrow \quad & v(x) = -\frac{ln(x)}{a}, \quad -\infty < x < \infty. \end{aligned}$$

The risk tolerance function is

$$\tau(x) = \frac{1}{a}, \quad -\infty < x < \infty. \tag{21}$$

Thus, (19) becomes

$$W_T = v(m\Psi) = \frac{-\big[ln(m) + ln(\Psi)\big]}{a}$$

and (20) becomes

$$E\big[\Psi v(m\Psi)\big] = \frac{-\big\{ln(m) + E\big[\Psi ln(\Psi)\big]\big\}}{a} = \omega e^{rT}.$$

By eliminating the $ln(m)$ term, we obtain

$$W_T = \omega e^{rT} + \frac{E\big[\Psi ln(\Psi)\big] - ln(\Psi)}{a}. \tag{22}$$

4.3.2 Decreasing LRT Utility Functions

Decreasing LRT utility functions are power utility functions with parameters c and s. $c > 0$ is the power of the marginal utility, meaning $-1/c$ is the slope of the risk tolerance function, and s is the finite level of maximal satisfaction:

$$\begin{aligned} u(x) &= -\frac{(s-x)^{c+1}}{c+1}, \quad x < s, \\ \Rightarrow \quad & u'(x) = (s-x)^c, \quad x < s, \\ \Leftrightarrow \quad & v(x) = s - x^{1/c}, \quad x < s. \end{aligned}$$

The risk tolerance function is

$$\tau(x) = \frac{s-x}{c}, \quad x < s. \tag{23}$$

Thus, (19) becomes

$$W_T = v(m\Psi) = s - m^{1/c}\Psi^{1/c}$$

and (20) becomes

$$E\big[\Psi v(m\Psi)\big] = s - m^{1/c}E\big[\Psi^{1+1/c}\big] = \omega e^{rT}.$$

By eliminating the $m^{1/c}$ term, we obtain

$$W_T = s - \frac{s - \omega e^{rT}}{E\big[\Psi^{1+1/c}\big]}\Psi^{1/c}, \quad \omega < se^{-rT}. \tag{24}$$

For $\omega \geq se^{-rT}$, the investor achieves maximal satisfaction by investing all of his money in the risk-free asset, which is the trivial solution for the investment problem.

4.3.3 Increasing LRT Utility Functions

Increasing LRT utility functions are power utility functions with parameters $c > 0$ and s. The value $-c$ is the power of the marginal utility, meaning $1/c$ is the slope of the risk tolerance function, and s is the minimal requirement of the terminal wealth:

$$u(x) = \begin{cases} \frac{(x-s)^{1-c}}{1-c}, & \text{for } x > s,\ c \neq 1, \\ ln(x-s), & \text{for } x > s,\ c = 1 \end{cases}$$
$$\Rightarrow \quad u'(x) = (x-s)^{-c}, \qquad x > s,$$
$$\Leftrightarrow \quad v(x) = s + x^{-1/c}, \qquad x > s.$$

The risk tolerance function is

$$\tau(x) = \frac{x-s}{c}, \qquad x > s. \tag{25}$$

Thus, (19) becomes

$$W_T = v(m\Psi) = s + m^{-1/c}\Psi^{-1/c}$$

and (20) becomes

$$E\big[\Psi v(m\Psi)\big] = s + m^{-1/c}E\big[\Psi^{1-1/c}\big] = \omega e^{rT}.$$

By eliminating the $m^{-1/c}$ term, we obtain

$$W_T = s + \frac{\omega e^{rT} - s}{E\big[\Psi^{1-1/c}\big]}\Psi^{-1/c}, \qquad \omega \geq se^{-rT}. \tag{26}$$

If $\omega < se^{-rT}$, the investment problem has no solution. The terminal wealth (11) will be below the minimal required wealth with positive probability for any random payment Y. Thus, the situation is hopeless for the investor.

4.3.4 Lognormal Price Density

We consider now the situation where the price density Ψ has a lognormal distribution, that is where $ln(\Psi)$ is normally distributed. Then the distribution of the optimal terminal wealth can be identified for the three cases of different LRT utility functions:

1. From (22) we see that W_T has a normal distribution.
2. From (24) it follows that W_T is the level of saturation minus a lognormal random variable.
3. From (26) we gather that W_T is the minimal required wealth level plus a lognormal random variable.

Since $S(T) = S(0)\ e^{X(T)}$, it follows from (2) that

$$\Psi = \frac{e^{h^* X(T)}}{E\left[e^{h^* X(T)}\right]}. \tag{27}$$

Using (5),

$$E\left[e^{h^* X(T)}\right] = exp\left[(h^* \mu + \frac{1}{2} h^{*2} \sigma^2) T\right]\ ,$$

and with the definition

$$\alpha = h^* \mu + \frac{(h^* \sigma)^2}{2}, \tag{28}$$

we get

$$\begin{aligned}
\Psi &= \frac{e^{h^* X(T)}}{e^{[h^*\mu T + \frac{1}{2}h^{*2}\sigma^2 T]}} \\
&= e^{[h^* X(T) - (h^*\mu + \frac{1}{2}h^{*2}\sigma^2)T]} \\
&= e^{[h^* X(T) - \alpha T]}. \qquad (29)
\end{aligned}$$

Since $X(T) \sim N(\mu T, \sigma^2 T)$, it follows that

$$E\big[ln(\Psi)\big] = E\big[h^* X(T) - \alpha T\big] = (h^*\mu - \alpha)T, \qquad (30)$$

and

$$\begin{aligned}
E\big[\Psi ln(\Psi)\big] &= E\big[\Psi h^* X(T) - \alpha T \Psi\big] \\
&= h^* E\big[X(T)\Psi\big] - \alpha T E\big[\Psi\big] \\
&= h^* E_Q\big[X(T)\big] - \alpha T \cdot 1 \\
&= h^* \mu^* T - \alpha T \qquad \text{by Remark 3} \\
&= (h^* \mu^* - \alpha)T, \qquad (31)
\end{aligned}$$

where μ^* is the risk-neutral drift of {X(T)} defined by (7).

We consider first the case of an exponential utility function, where the optimal terminal wealth is normally distributed. Therefore, (22) changes to

$$\begin{aligned}
W_T &= \omega e^{rT} + \frac{1}{a}\big(E[\Psi ln(\Psi)] - ln(\Psi)\big) \\
&= \omega e^{rT} + \frac{1}{a}\big((h^*\mu^* - \alpha)T - ln(\Psi)\big) \quad \text{using (31)}.
\end{aligned}$$

Then,

$$\begin{aligned} E[W_T] &= \omega e^{rT} + \frac{h^*\mu^*T - \alpha T - E[ln(\Psi)]}{a} \\ &= \omega e^{rT} + \frac{h^*\mu^*T - \alpha T - h^*\mu T + \alpha T}{a} \\ &= \omega e^{rT} + \frac{h^*T(\mu^* - \mu)}{a} \\ &= \omega e^{rT} + \frac{(h^*\sigma)^2 T}{a} \quad \text{by (8)} \end{aligned} \tag{32}$$

and

$$\begin{aligned} Var[W_T] &= \frac{Var(ln(\Psi))}{a^2} \\ &= \frac{Var(h^*X(T) - \alpha T)}{a^2} \quad \text{by (29)} \\ &= \frac{h^{*2}Var(X(T))}{a^2} \\ &= \frac{(h^*\sigma)^2 T}{a^2}. \end{aligned} \tag{33}$$

Thus, the higher the risk tolerance ($1/a$), the higher is the expectation of terminal wealth, but with the effect of a progressively increasing variance.

Next, we consider the case of an increasing LRT utility function. We use for the calculation of the expected value and the variance of the optimal terminal wealth the following result of probability calculus:

Theorem 4 *For a normal random variable Z, we have*

$$\begin{aligned} E\left[e^Z\right] &= exp\left(E[Z] + \frac{1}{2}Var[Z]\right) \\ Var\left[e^Z\right] &= e^{2E[Z]+Var[Z]}\left(e^{Var[Z]} - 1\right). \end{aligned}$$

Proof:

see Aitchison and Brown (1957).

□

We have

$$E[\Psi^{1-1/c}] = E\left[e^{(h^*X(T)-\alpha T)(1-1/c)}\right] \qquad \text{by (29)}$$

$$= E\left[e^{h^*X(T)-\alpha T-h^*X(T)/c+\alpha T/c}\right]$$

$$= e^{\alpha T(\frac{1}{c}-1)} \cdot E[e^{X(T)(h^*-h^*/c)}]$$

$$= e^{\left[\alpha T(\frac{1}{c}-1)\right]} \cdot e^{\left[E[X(T)(h^*-h^*/c)]+\frac{1}{2}Var[X(T)(h^*-h^*/c)]\right]} \quad \text{by Theorem 4}$$

$$= e^{\left[h^*\mu T(\frac{1}{c}-1)+\frac{(h^*\sigma)^2}{2}T(\frac{1}{c}-1)\right]} \cdot e^{\left[\mu T h^*(1-1/c)+\frac{1}{2}h^{*2}(1-1/c)^2\sigma^2 T\right]}$$

$$= e^{(h^*\sigma)^2 T\left(\frac{1}{2c}-\frac{1}{2}+\frac{1}{2}(1-2/c+1/c^2)\right)}$$

$$= e^{(h^*\sigma)^2 T\left(-\frac{1}{2c}+\frac{1}{2c^2}\right)}.$$

Hence (26) becomes

$$\begin{aligned}
W_T &= s + (\omega e^{rT} - s)\Psi^{-1/c} e^{-(h^*\sigma)^2 T\left(-\frac{1}{2c}+\frac{1}{2c^2}\right)} \\
&= s + (\omega e^{rT} - s)\ exp\left[-h^*X(T)/c + \alpha T/c - (h^*\sigma)^2 T\left(-\frac{1}{2c}+\frac{1}{2c^2}\right)\right] \text{ by (29)} \\
&= s + (\omega e^{rT} - s)\ exp\left[-h^*X(T)/c + h^*\mu T/c + \frac{(h^*\sigma)^2}{2c}T + (h^*\sigma)^2 T\left(\frac{1}{2c}-\frac{1}{2c^2}\right)\right] \\
&= s + (\omega e^{rT} - s)\ exp\left[T\left\{(h^*\sigma)^2(\frac{1}{c}-\frac{1}{2c^2}) + \frac{h^*\mu}{c}\right\} - X(T)h^*/c\right] \\
&= s + (\omega e^{rT} - s)\ exp\left[\gamma T - (h^*/c)X(T)\right], \quad \omega \geq se^{-rT}, \qquad (34)
\end{aligned}$$

where

$$\gamma = (h^*\sigma)^2\left(\frac{1}{c}-\frac{1}{2c^2}\right) + \frac{h^*\mu}{c}. \qquad (35)$$

Thus

$$W_T = \begin{cases} \omega e^{rT} & \text{if } \omega = se^{-rT} \\ s + e^Z & \text{if } \omega > se^{-rT} \end{cases} \tag{36}$$

where

$$Z = ln(\omega e^{rT} - s) + \gamma T - \frac{h^* X(T)}{c}$$

is a normal random variable with

$$\begin{aligned} E[Z] &= ln(\omega e^{rT} - s) + (h^*\sigma)^2 \left(\frac{1}{c} - \frac{1}{2c^2}\right) T + \frac{h^*\mu}{c}\, T - \frac{h^*}{c}\, E[X(T)] \\ &= ln(\omega e^{rT} - s) + (h^*\sigma)^2 \left(\frac{1}{c} - \frac{1}{2c^2}\right) T + \frac{h^*\mu}{c}\, T - \frac{h^*}{c}\, \mu T \\ &= ln(\omega e^{rT} - s) + (h^*\sigma)^2 \left(\frac{1}{c} - \frac{1}{2c^2}\right) T \\ Var[Z] &= \left(\frac{h^*}{c}\right)^2 Var[X(T)] \\ &= \left(\frac{h^*\sigma}{c}\right)^2 T \end{aligned}$$

By using the results and Theorem 4, we can calculate the expected value and variance of the optimal terminal wealth for the case of an increasing LRT utility function:

$$\begin{aligned} E[W_T] &= E\big[s + e^Z\big] = s + exp\left(E[Z] + \frac{1}{2} Var[Z]\right) \\ &= s + exp\left(ln(\omega e^{rT} - s) + (h^*\sigma)^2 \left(\frac{1}{c} - \frac{1}{2c^2}\right) T + \frac{1}{2}\left(\frac{h^*\sigma}{c}\right)^2 T\right) \\ &= s + (\omega e^{rT} - s) exp\left[\frac{(h^*\sigma)^2 T}{c}\right], \quad \omega > se^{-rT} \end{aligned} \tag{37}$$

$$
\begin{aligned}
Var[W_T] &= Var\left[s+e^Z\right] = Var[e^Z] \\
&= exp\left[2E(Z)+Var(Z)\right]\cdot(e^{Var(Z)}-1) \\
&= exp\left[2ln(\omega e^{rT}-s)+2(h^*\sigma)^2\left(\frac{1}{c}-\frac{1}{2c^2}\right)T+\left(\frac{h^*\sigma}{c}\right)^2T\right]\cdot \\
&\quad \cdot\left(exp\left[\left(\frac{h^*\sigma}{c}\right)^2T\right]-1\right) \\
&= exp\left[\left(\frac{h^*\sigma}{c}\right)^2T+2ln(\omega e^{rT}-s)+2(h^*\sigma)^2\left(\frac{1}{c}-\frac{1}{2c^2}\right)T+\left(\frac{h^*\sigma}{c}\right)^2T\right] \\
&\quad -exp\left[2ln(\omega e^{rT}-s)+2(h^*\sigma)^2\left(\frac{1}{c}-\frac{1}{2c^2}\right)T+\left(\frac{h^*\sigma}{c}\right)^2T\right] \\
&= (\omega e^{rT}-s)^2\Bigg\{exp\left[2\left(\frac{h^*\sigma}{c}\right)^2T+2\frac{(h^*\sigma)^2}{c}T-\left(\frac{h^*\sigma}{c}\right)^2T\right] \\
&\qquad\qquad -exp\left[2\frac{(h^*\sigma)^2}{c}T\right]\Bigg\} \\
&= (\omega e^{rT}-s)^2\left\{exp\left[\left(\frac{h^*\sigma}{c}\right)^2T+2\frac{(h^*\sigma)^2}{c}T\right]-exp\left[2\frac{(h^*\sigma)^2}{c}T\right]\right\} \qquad (38)
\end{aligned}
$$

Gerber and Shiu (2000) state as solution:

$$
Var[W_T] = (\omega e^{rT}-s)^2\left[exp\left(\frac{3(h^*\sigma)^2T}{c}\right)-exp\left(\frac{2(h^*\sigma)^2T}{c}\right)\right], \qquad \omega > se^{-rT}.
$$

It follows from (37) that, as the minimal required wealth level s increases from 0 to ωe^{rT}, the mean of the optimal terminal wealth, $E[W_T]$, decreases linearly to ωe^{rT}, and the variance, $Var[W_T]$, decreases quadratically to zero. Likewise, the mean and the variance of W_T are decreasing functions of the parameter c. These properties recall that the risk tolerance function $\tau(x)$, as given in (25), is a decreasing function in each of the parameters s and c.

For $0 < p < 1$, let ζ_p denote the $100p$-th percentile of the standard normal distribution, that is $p = \int_{-\infty}^{\zeta_p} f(x)dx = \Phi(\zeta_p)$, where $f(x)$ is the p.d.f. of the standard normal distribution. Gerber and Shiu (2000) show that the 100p-th percentile of the distribution of W_T is

$$s + exp\left(E[Z] + \zeta_p \sqrt{Var[Z]}\right), \quad \omega > se^{-rT} , \tag{39}$$

and the median of the distribution of W_T is

$$\begin{aligned} \text{Median}(W_T) &= Mdn(W_T) = s + e^{E[Z]} \\ &= s + (\omega e^{rT} - s)\ exp\left[(h^*\sigma)^2 T\left(\frac{1}{c} - \frac{1}{2c^2}\right)\right], \quad \omega > se^{-rT}. \end{aligned} \tag{40}$$

For the calculation of the Mode(W_T), we apply the formula for the mode of a lognormal distribution (Aitchison and Brown, 1957) to (36) to get

$$\begin{aligned} \text{Mode}(W_T) &= s + e^{E[Z] - Var[Z]} \qquad (41) \\ &= s + exp\left(ln(\omega e^{rT} - s) + (h^*\sigma)^2\left(\frac{1}{c} - \frac{1}{2c^2}\right)T - \left(\frac{h^*\sigma}{c}\right)^2 T\right) \\ &= s + (\omega e^{rT} - s) exp\left[(h^*\sigma)^2 T\left(\frac{1}{c} - \frac{3}{2c^2}\right)\right], \quad \omega > se^{-rT}. \qquad (42) \end{aligned}$$

Comparing (42) with (37) and (40), we see that for $\omega > se^{-rT}$,

Mode(W_T) < Mdn(W_T) < ωe^{rT} < $E[W_T]$, $0 < c < \frac{1}{2}$

Mode(W_T) < ωe^{rT} < Mdn(W_T) < $E[W_T]$, $\frac{1}{2} < c < \frac{3}{2}$

ωe^{rT} <Mode(W_T) < Mdn(W_T) < $E[W_T]$, $\frac{3}{2} < c$.

As s increases to ωe^{rT}, we have the following impacts:

- W_T converges to the degenerate random variable with the constant value ωe^{rT}
- the mean of W_T decreases linearly as mentioned earlier in this chapter

- the median of W_T $\begin{cases} \text{increases, if } c < \frac{1}{2} \\ \text{decreases, if } c > \frac{1}{2} \end{cases}$
- the mode of W_T $\begin{cases} \text{decreases, if } c < \frac{3}{2} \\ \text{increases, if } c > \frac{3}{2} \end{cases}$

The mean of W_T is also a decreasing function of c, which is in contrast to the median and mode of W_T. After differentiating (40) and (42) with respect to c, we can easily see that the median attains its maximum for $c = 1$, while the mode has its maximum for $c = 3$.

Remark 6 *The value $(h^*\sigma)^2$ appears frequently in this section and is the square of the "market price of risk".*

If $c = -h^* > 0$, then (35) changes to

$$\begin{aligned} \gamma &= (h^*\sigma)^2 \left(-\frac{1}{h^*} - \frac{1}{2(h^*)^2} \right) - \mu \\ &= -h^*\sigma^2 - \frac{\sigma^2}{2} - \mu \\ &= -(r - \frac{\sigma^2}{2} - \mu) - \frac{\sigma^2}{2} - \mu \quad \text{by (6)} \\ &= -r \end{aligned}$$

and hence (34) becomes

$$\begin{aligned} W_T &= s + \left(\omega e^{rT} - s\right) e^{-rT+X(T)} \\ &= s + \left(\omega - se^{-rT}\right) e^{X(T)}, \end{aligned} \tag{43}$$

meaning that an investor with this particular utility function can obtain his optimal terminal wealth by using the following strategy: investing the amount se^{-rT} in the risk-free asset and the remaining part, $\omega - se^{-rT}$, in the risky asset at time 0. After having introduced dynamic investment strategies, we will return to this strategy, which is also called "buy-and-hold" investment strategy.

4.4 Utility of the Initial Wealth

The *maximal expected utility* of terminal wealth is a function of the initial wealth ω and we denote it as

$$u_0(\omega) = E\big[u(W_T)\big]. \tag{44}$$

In the actuarial literature, $u_0(\omega)$ is also called a derived utility function, an implied utility function, an indirect utility function, or an induced utility function. We will prove the following relation, that shows the connection of the maximal expected utility with the function m defined by (16):

$$u_0'(\omega) = me^{rT}. \tag{45}$$

Proof: We rewrite (44) by using (19)

$$u_0(\omega) = E[u(v(m\Psi))].$$

Applying the chain rule, we get

$$\begin{aligned} u_0'(\omega) &= \frac{d}{dm}E\big[u(v(m\Psi))\big]\frac{dm}{d\omega} \\ &= E\big[u'(W_T)v'(m\Psi)\Psi\big]\Big/\frac{d\omega}{dm} \\ &= mE\big[v'(m\Psi)\Psi^2\big]\Big/\frac{d\omega}{dm} \qquad \text{by (17).} \end{aligned}$$

Then we use the result from (20)

$$\frac{d\omega}{dm} = e^{-rT} E\big[v'(m\Psi)\Psi^2\big] \tag{46}$$

to obtain

$$\begin{aligned} u_0'(\omega) &= mE\big[v'(m\Psi)\Psi^2\big]/e^{-rT}E\big[v'(m\Psi)\Psi^2\big] \\ &= me^{rT}. \end{aligned}$$

□

Looking at the equivalent notation of (45), $\frac{d}{d\omega}E[u(W_T)] = E[u'(W_T)]e^{rT}$, one might pose the question, if there is also an equality between the random variables $d/d\omega\ u(W_T)$ and $u'(W_T)e^{rT}$, or equivalently if $d/d\omega\ W_T$ can be e^{rT}. This is true for the exponential utility functions, but it is not valid for the power utility functions, which can easily be seen by inspecting (22), (24), and (26).

The maximal expected utility function $u_0(\omega)$ has the two properties of a risk-averse utility function:

1. $u_0' > 0$:

   (45) and (16) imply that the first derivative u_0' has the same sign as u', which is positive.

2. $u_0'' < 0$:

   Taking the derivative of (45), we get

$$u_0''(\omega) = e^{rT}\frac{dm}{d\omega} = e^{rT} \Big/ \frac{d\omega}{dm}. \tag{47}$$

Applying (46) yields

$$u_0''(\omega) = \frac{e^{2rT}}{E\big[v'(m\Psi)\Psi^2\big]}.$$

By taking the derivative of (18), we get $v'(u'(x))u''(x) = 1$. Since u'' is negative, it follows that v' is negative. Hence, u_0'' is negative.

Let

$$\tau_0(\omega) = -\frac{u_0'(\omega)}{u_0''(\omega)} \tag{48}$$

be the risk tolerance function that is associated with the induced utility function $u_0(\omega)$. (45) and (47) imply that

$$\tau_0(\omega) = -m\frac{d\omega}{dm}. \tag{49}$$

We will use this result in the next section.

We can show that the following relation is true for LRT utility functions:

$$\tau_0(\omega) = e^{-rT}\tau(\omega e^{rT}), \tag{50}$$

or alternatively, that $u_0(\omega)$ is equivalent to $u(\omega e^{rT})$.

CHAPTER V

OPTIMAL DYNAMIC INVESTMENT STRATEGIES

In the previous chapter we considered a one-period securities market model in which random payments due at time T can be traded at time 0. If such a market does not exist, how can we create the random payments? It may be possible to create them in a synthetic way by dynamically trading the primitive securities. We consider a market in which securities can be traded at all times t, $0 \leq t \leq T$ and we assume that the market is frictionless and there are no taxes. Therefore, tradings will not add any cost. In a *complete* securities market, each contingent claim or random payment can be replicated by a self-financing portfolio of marketable securities under the assumption that the market is arbitrage-free.

Our goal is to replicate the optimal terminal wealth W_T in such a securities market to get the optimal dynamic investment strategy, starting with the amount ω at time 0.

5.1 A Single Risky Asset

We start the investigation with a single risky asset in this section and we transfer the results to the dynamic market model with multiple risky assets in the next section.

The model considered is a complete securities market model. (19) and (29) provide the formula for the optimal terminal wealth

$$W_T = v(m\Psi) = v(me^{h^* X(T) - \alpha T}). \tag{51}$$

As in Panjer (1998), we consider a contingent claim at time T, which is a function of the risky asset price at time T, $S(T)$. But it does not depend on the asset prices before time T. In terms of the *payoff* function, $\pi(\cdot)$, the contingent payment is $\pi(S(T))$. We denote $V(s,t)$ the price for the contingent claim at time t, given that $s = S(t),\ 0 \le t < T$:

$$V(s,t) = e^{-r(T-t)} E\big[\pi(S(T))\Psi | s = S(t)\big]. \tag{52}$$

Since the market is complete, the contingent claim can be replicated by a dynamic, self-financing portfolio. Let $\eta(S(t),t)$ denote the amount in the replicating portfolio invested in the risky asset at time $t, 0 \le t \le T$. Hence, the amount $V(S(t),t) - \eta(S(t),t)$ is invested in the risk-free asset. Gerber and Shiu (1996) show that

$$\eta(s,t) = s\frac{\partial}{\partial s}V(s,t). \tag{53}$$

The second part of the product, the partial derivative $\frac{\partial}{\partial s}V(s,t)$, is also called *delta* in the option-pricing literature.

To apply formula (53), consider a related contingent claim defined by

$$\begin{aligned} \pi(S(T)) &= v(me^{-\alpha T}S(T)^{h^*}) \\ &= v(mS(0)^{h^*}\Psi). \end{aligned} \tag{54}$$

Using the rewritten formula (20)

$$\omega(m) = \omega = e^{-rT}E[\Psi v(m\Psi)],$$

we can determine the time-0 price of (54):

$$\begin{aligned} V(S(0),0) &= e^{-rT}E\big[\pi(S(T))\Psi|s=S(0)\big] \\ &= e^{-rT}E\big[v(mS(0)^{h^*}\Psi)\Psi|s=S(0)\big] \\ &= \omega(mS(0)^{h^*}). \end{aligned} \tag{55}$$

Applying the chain rule to (55), we get from (53),

$$\begin{aligned} \eta(s,0) &= s\frac{\partial}{\partial s}V(s,0) = s\frac{\partial}{\partial s}\omega(ms^{h^*}) \\ &= s\omega'(ms^{h^*})mh^*s^{h^*-1} \\ &= h^*ms^{h^*}\omega'(ms^{h^*}) \end{aligned} \tag{56}$$

for the replicating portfolio of (54).

We are still looking for the replicating portfolio of W_T. Let W_t denote the value of the replicating portfolio at time t, $0 \le t < T$. In other words, the investor's initial investment of $W_0 = \omega$ accumulates to wealth W_t at time t by using the optimal strategy.

We denominate the amount in the replicating portfolio invested in the risky asset at time t, $\rho(W_t, t)$. Likewise, $W_t - \rho(W_t, t)$ is the amount invested in the risk-free asset. If we set $S(0) = 1$, formulas (54) and (51) are equal. Hence,

$$\rho(\omega,0) = \eta(1,0)$$

$$= h^*m\,\omega'(m) \qquad \text{by (56)} \tag{57}$$

$$= -h^*\tau_0(\omega) \qquad \text{by (49)} \tag{58}$$

$$= \frac{\mu-\mu^*}{\sigma^2}\,\tau_0(\omega). \qquad \text{by (8)} \tag{59}$$

To extend these formulas to time $t, 0 < t < T$, let $u_t(y)$ denote the conditional expected utility of optimal terminal wealth, given that $W_t = y$. Furthermore, let $\tau_t(\cdot)$ be the corresponding risk tolerance function. Hence the random variable $\tau_t(W_t)$ is the implied risk tolerance at time t, and we call it current risk tolerance. We generalize (58) and (59) to obtain an investment rule:

$$\rho(W_t, t) = -h^* \, \tau_t(W_t) \tag{60}$$

$$= \frac{\mu - \mu^*}{\sigma^2} \, \tau_t(W_t), \qquad 0 \leq t < T. \tag{61}$$

The investment rule (61) determines the optimal amount invested in the risky asset as the product of the current risk tolerance and the risk premium on the risky asset, divided by the square of the diffusion coefficient.

To discuss the optimal investment in the risky asset as a fraction of the investor's current wealth at time t, we introduce a general version of the *Merton ratio*, in honor of the Nobel laureate Robert C. Merton:

$$M(W_t, t) = \frac{\rho(W_t, t)}{W_t}. \tag{62}$$

Using (57) we see that

$$M(\omega, 0) = \frac{h^* m \, \omega'(m)}{\omega}. \tag{63}$$

We will come back to this formula and restate it in the next chapter, after introducing the concept of elasticity.

It turns out that for LRT utility functions we can obtain some very explicit formulas for the Merton ratio: Using (50), we can rewrite (58) and (59) as

$$\begin{aligned} \rho(\omega,0) &= -h^* e^{-rT} \tau(\omega e^{rT}) \\ &= \frac{\mu-\mu^*}{\sigma^2}\, e^{-rT} \tau(\omega e^{rT}). \end{aligned} \tag{64}$$

We substitute this in (62) and get

$$\begin{aligned} M(\omega,0) &= -h^* e^{-rT} \tau(\omega e^{rT})/\omega \\ &= \frac{\mu-\mu^*}{\sigma^2\omega}\, e^{-rT} \tau(\omega e^{rT}). \end{aligned} \tag{65}$$

5.2 Constant Risk Tolerance Utility Functions

For constant risk tolerance utility functions, it follows from (21) that

$$\rho(\omega,0) = \frac{-h^* e^{-rT}}{a} \tag{66}$$

and

$$M(\omega,0) = \frac{-h^* e^{-rT}}{a\omega}. \tag{67}$$

For $0 \leq t < T$, we obtain from (66)

$$\begin{aligned} \rho(W_t,t) &= \frac{-h^* e^{-r(T-t)}}{a} \\ &= \frac{\mu-\mu^*}{\sigma^2 a}\, e^{-r(T-t)}, \end{aligned} \tag{68}$$

which can be interpreted as follows: for an exponential utility function, the amount invested in the risky asset accumulated with interest is the constant $-h^*/a$. The reason that the amount invested is independent of the investor's wealth reflects the fact that his risk tolerance function is constant.

5.3 Decreasing LRT Utility Functions

For decreasing LRT utility functions, it follows from (23) that

$$\rho(\omega, 0) = -(h^*/c)(se^{-rT} - \omega), \quad \omega < se^{-rT} \tag{69}$$

and

$$M(\omega, 0) = \frac{-(h^*/c)(se^{-rT} - \omega)}{\omega}, \quad \omega < se^{-rT}. \tag{70}$$

We generalize (70) for $0 \leq t < T$ to obtain

$$\begin{aligned} M(W_t, t) &= \frac{-(h^*/c)(se^{-r(T-t)} - W_t)}{W_t} \\ &= \frac{\mu - \mu^*}{\sigma^2 c} \frac{se^{-r(T-t)} - W_t}{W_t}, \quad W_t < se^{-r(T-t)}, \end{aligned} \tag{71}$$

which means that the optimal amount invested in the risky asset is proportional to what the current wealth is less than the discounted value of the minimal required terminal wealth.

5.4 Increasing LRT Utility Functions

For increasing LRT utility functions, it follows from (25) that

$$\rho(\omega, 0) = -(h^*/c)(\omega - se^{-rT}), \quad \omega \geq se^{-rT} \tag{72}$$

and

$$M(\omega, 0) = \frac{-(h^*/c)(\omega - se^{-rT})}{\omega}, \qquad \omega \geq se^{-rT}. \tag{73}$$

Generalizing (73) for $0 \leq t < T$, yields to

$$\begin{aligned} M(W_t, t) &= \frac{-(h^*/c)(W_t - se^{-r(T-t)})}{W_t} \\ &= \frac{\mu - \mu^*}{\sigma^2 c} \, \frac{W_t - se^{-r(T-t)}}{W_t}, \qquad W_t \geq se^{-r(T-t)}, \end{aligned} \tag{74}$$

which says that the optimal amount invested in the risky asset is proportional to the excess of current wealth over the discounted level of saturation.

An investor may have a terminal wealth of at least $s, 0 < s \leq \omega e^{rT}$. Then the portion of current wealth W_t that grows to s at time T with certainty is $se^{-r(T-t)}$. The remaining part of the current wealth, $W_t - se^{-r(T-t)}$, can be considered "free". Hence (74) has the meaning that at any time, a constant portion of the "free" wealth should be invested in the risky asset.

In the special case $s = 0$, we get from (74) the portion of the current wealth that should be invested in the risky asset:

$$M(W_t, t) = \frac{-h^*}{c} = \frac{\mu - \mu^*}{\sigma^2 c}, \tag{75}$$

which is constant. This formula is usually called the Merton ratio in the literature (Panjer, 1998), not the more general formula (62).

Example:

We consider an investor who has an increasing LRT utility function with $c = 4/3$ and $s = 0$. Let the risk premium on the risky asset, $\mu - \mu^*$ be 5% and let the volatility of the risky asset as measured by σ be 25%. Then the Merton ratio (75) for this investor is 60%, which is the figure in the rule of thumb mentioned in the introduction.

5.5 Optimality of Buy-and-Hold Strategies

As a result of formula (43), we know that an investor who has an increasing LRT utility function with $c = -h^*$ uses a buy-and-hold investment strategy. Now we show that the converse is also true:

We assume that the buy-and-hold investment strategy is optimal for an investor, who invests the amount $a, a \leq \omega$, in the risk-free asset and the complement, $\omega - a$, in the risky asset at time 0. Since the investor uses the buy-and-hold investment strategy, there are no rebalances in the portfolio. Hence,

$$W_t = ae^{rt} + \rho(W_t, t), \quad 0 \leq t < T. \tag{76}$$

Because of (76), (60) changes to

$$W_t - ae^{rt} = -h^* \; \tau_t(W_t), \quad 0 \leq t < T, \tag{77}$$

and taking the limit $t \to T$, we get

$$W_T - ae^{rT} = -h^* \; \tau(W_T)$$

$$\Leftrightarrow \tau(W_T) = \frac{W_T \; - \; ae^{rT}}{-h^*}$$

Since the amount a is invested in the risk-free asset, W_T can take on any value between ae^{rT} and ∞. Thus we have

$$\tau(x) = \frac{x - ae^{rT}}{-h^*} \;\; \text{for } x \geq ae^{rT}, \tag{78}$$

which is the risk tolerance function of an increasing LRT utility function with $s = ae^{rT}$ and $c = -h^*$ as stated in (25). Hence, the investor has an increasing LRT utility function.

Remark 7 *The property of the buy-and-hold investment strategy can also be viewed in a more general framework. In the last three sections we arrived at the conclusion that for LRT utility functions the optimal amount invested in the risky asset at any time t is a linear function of the investor's current wealth W_t. Again, the converse also holds. We assume that the optimal amount invested in the risky asset at any time is a linear function of the investor's current wealth, that is,*

$$\rho(W_t, t) = a(t) + b(t)W_t, \quad 0 \leq t < T,$$

where $a(t)$ and $b(t)$ are functions with existing limits for $t \to T$, $a(T)$ and $b(T)$. With (60), we get

$$\tau(x) = \frac{a(T) + b(T)x}{-h},$$

which means that the investor has a LRT utility function.

5.6 Multiple Risky Assets

We generalize now the optimal dynamic investment strategies to the case of multiple primitive assets. In this model, the price density is

$$\Psi = e^{-\alpha T} \sum_{k=1}^{n} \left[\frac{S_k(T)}{S_k(0)} \right]^{h_k^*}, \tag{79}$$

where α is such that $E[\Psi] = 1$. Again, we have following notations:

- $W_0 = \omega$

- W_t is the value of the replicating portfolio for W_T, $0 < t < T$

- $\rho_k(W_t, t)$ is the amount invested in risky asset k in the replicating portfolio of W_T at time t, $0 \leq t < T$ and $k = 1, 2, ..., n$

For $k = 1, 2, ..., n$ let

$$M_k(W_t, t) = \frac{\rho_k(W_t, t)}{W_t}$$

be the Merton ratio for risky asset k at time t, which generalizes (62). To get the corresponding formula for (53), we consider a contingent claim with payoff

$$\pi(S_1(T), S_2(T), ..., S_n(T)) \tag{80}$$

at time T for a payoff function π. With the notation

- $V(s_1, s_2, ..., s_n, t)$ = price of the contingent claim at time t

- For $k = 1, 2, ..., n$, $\eta_k(s_1, s_2, ..., s_n, t)$ is the amount of risky asset k in the replicating portfolio for (80), given that $S_j(t) = s_j$, $j = 1, 2, ..., n$

Gerber and Shiu (1996) state that

$$\eta_k = s_k \frac{\partial V}{\partial s_k}. \tag{81}$$

We consider the payoff

$$\begin{aligned} \pi(S_1(T), S_2(T), ..., S_n(T)) &= v\left(me^{-\alpha T}\prod_{j=1}^{n} S_j(T)^{h_j^*}\right) \\ &= v\left(m\left[\prod_{j=1}^{n} S_j(0)^{h_j^*}\right]\Psi\right). \end{aligned} \tag{82}$$

The time-0 price of $W_T = v(m\Psi)$ is $\omega = \omega(m)$. Therefore, the time-0 price of (82) is

$$V(s_1, s_2, ..., s_n, t) = \omega\left(m\prod_{j=1}^{n} s_j^{h_j^*}\right), \tag{83}$$

where $s_j = S_j(0), j = 1, 2, ..., n$. Applying the chain rule to (83), we get

$$\eta_k(s_1, s_2, ..., s_n, 0) = \omega'\left(m\prod_{j=1}^{n} s_j^{h_j^*}\right) m\ h_k^* \prod_{j=1}^{n} s_j^{h_j^*}. \tag{84}$$

Then,

$$\begin{aligned}\rho_k(\omega, 0) &= \eta_k(1, 1, ..., 1, 0) \\ &= \omega'(m)\ m\ h_k^* \\ &= -h_k^*\ \tau_0(\omega) \quad \text{by (49).} \end{aligned} \tag{85}$$

Hence

$$\rho_k(W_t, t) = -h_k^*\ \tau_t(W_t), \tag{86}$$

which generalizes (61).

The amount invested in risky asset k as a fraction of the total amount invested in all risky assets depends only on the risk-neutral Esscher parameters and remains constant, say q_k at all times:

$$\frac{\rho_k(W_t, t)}{\sum_{j=1}^{n} \rho_j(W_t, t)} = \frac{h_k^*}{\sum_{j=1}^{n} h_j^*} = q_k. \tag{87}$$

Gerber and Shiu (2000) state this result as the "mutual fund" theorem: for any risk-averse investor, the risky-asset portion of his optimal investment portfolio is of identical composition. Thus it is sufficient to have two mutual funds, one risk-free bond fund, and one risky-asset mutual fund with the characteristic that at all times the fraction of its value invested in risky asset k is q_k. Each investor simply needs to invest, or to borrow from the two mutual funds. The amount of investment in the risky-asset mutual fund divided by the investor's current risk tolerance is always at the constant level $-\sum_{j=1}^{n} h_j^*$.

Remark 8 *(Gerber and Shiu, 2000) The continuously compounded rate of return over the time interval* $(0,t)$ *of the risky-asset mutual fund is*

$$X(t) = \sum_{k=1}^{n} q_k \ X_k(t).$$

The stochastic process $\{X(t)\}$ *is a Wiener process with drift and diffusion parameters*

$$\begin{aligned} \mu &= \sum_{k=1}^{n} q_k \ \mu_k \\ \sigma^2 &= \sum_{i=1}^{n} \sum_{j=1}^{n} q_i \ \sigma_{ij} \ q_j \ . \end{aligned}$$

5.7 Elasticity

The following is a standard concept used in economics, and utilizes by Gerber and Shiu (2000).

Definition 16 *Let* $: \mathbb{R} \to \mathbb{R}$ *be a differentiable function,* $y = f(x)$*, with derivative* $y' = dy/dx = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$*. The limit of the relative changes,*

$$\lim_{\Delta x \to 0} \left(\frac{\Delta y}{y} \right) \Big/ \left(\frac{\Delta x}{x} \right) \tag{88}$$

is called the elasticity of y *with respect to* x *and denoted by the symbol* $\epsilon_x(y)$.

From (88), we get

$$\epsilon_x(y) = \frac{xy'}{y} = \frac{y'}{y/x} = x(lny)' \ . \tag{89}$$

As an interpretation, we concentrate on the following representation:

$$\epsilon_x(y) = \frac{y'}{y/x} \ , \tag{90}$$

which is the ratio of the slope of the tangent at (x, y) divided by the slope of the line that connects the origin with (x, y). By interchanging the roles of x and y in (88), we see that the elasticity of the inverse is the reciprocal of the original elasticity:

$$\epsilon_y(x) = \frac{1}{\epsilon_x(y)} \ . \tag{91}$$

We apply the concept of elasticity to the formula of the Merton ratio (63), which is the product of the risk-neutral Esscher parameter h^* with the elasticity of the initial wealth ω with respect to the expected marginal utility of optimal terminal wealth m,

$$\epsilon_m(\omega) = \frac{m \ \omega'(m)}{\omega} \ . \tag{92}$$

Therefore, (63) changes to

$$M(\omega, 0) = h^* \epsilon_m(\omega) = \frac{h^*}{\epsilon_\omega(m)}, \tag{93}$$

and for multiple risky assets we get

$$M_k(\omega, 0) = h_k^* \epsilon_m(\omega) = \frac{h_k^*}{\epsilon_\omega(m)}, k = 1, 2, ..., n. \tag{94}$$

For the LRT utility functions, we can give explicit expressions for the elasticities.

- For the constant risk tolerance utility functions, we have from (93) and (67)

$$\epsilon_\omega(m) = -a\ \omega\ e^{rT}.$$

- For the decreasing LRT utility functions, we obtain from (93) and (70)

$$\epsilon_\omega(m) = \frac{-c\ \omega}{se^{rT} - \omega}, \qquad \omega < se^{-rT}.$$

- For the increasing LRT utility functions, we find from (93) and (73)

$$\epsilon_\omega(m) = \frac{-c\ \omega}{\omega - se^{rT}}, \qquad \omega > se^{-rT}.$$

  In the special case $s = 0$, we obtain

$$\epsilon_\omega(m) = -c.$$

REFERENCES

Allen, Everett, T. et al. 2003. *Pension Planning: Pension, Profit Sharing, and Other Deferred Compensation Plans.* 9th ed. New York, NY: McGraw-Hill/Irwin.

Black, Kenneth, Jr. 1955. *Group Annuities.* Philadelphia, PH: University of Pennsylvania Press.

Bluhm, William, F. 2000. *Group Insurance.* 3rd ed. Winsted, Conn.: ACTEX Publications.

Bodie, Z., Kane, A., and Marcus, A. J. 1999. *Investments.* 4th ed. McGraw-Hill Companies.

Boersch-Supan, Axel 2000. "A Model Under Siege: A Case Study of the German Retirement Insurance System." *The Economic Journal* 110: F24-F45.

Bonin, Holger 2001. "Will it Last? An Assessment of the 2001 German Pension Reform." IZA Discussion Paper No. 343, Institute for the Study of Labor, Bonn, Germany.

Bos, Eduard, R. et al. 1994. *World Population Projections, 1994-95.* Washington D.C.: The International Bank for Reconstruction and Development / The World Bank.

Carpenter, Jill, K. 2000. "Introduction to Financial Security Systems." *Course 5 Study Note.* Schaumburg, IL: Society of Actuaries.

Department of Labor's Pension and Welfare Benefits Administration 2003. "Top 10 ways to beat the clock and prepare for retirement," Online at http://www.asec.org/tools/topten.htm.

Deutsche Bank Research 2003. "On the statutory pension scheme reform," Online at http://www.dbresearch.de/PROD/999/PROD0000000000060093.pdf.

Dothan, Michael, U. 1990. *Prices in Financial Markets.* New York, NY: Oxford University Press.

Employee benefit research institute (EBRI) 1997. *EBRI Databook on Employee Benefits.* Washington, DC.

Gajek, Leslaw and Ostaszewski, Krzysztof 2002. *Plany Emerytalne: Zarzadzanie Aktywami i Zobowiazaniami.* Wydawnictowa Naukowo-Techniczne, Warsaw, Poland.

Gajek, Leslaw and Ostaszewski, Krzysztof upcoming in 2003. *Pension Plans and Their Asset-Liability Management.* Boston, Massachusetts: Academic Press.

Gerber, Hans U. and Shiu, Elias, S.W. 2000. "Investing for Retirement: Optimal Capital Growth and Dynamic Asset Allocation," *North American Journal* 4 (2): 42-58.

Gerber, Hans U. and Shiu, Elias, S.W. 1988. "Utility Functions: From Risk Theory to Finance," *North American Journal* 2 (3): 74-100.

Gerber, Hans U. and Shiu, Elias, S.W. 1994. "Option Pricing by Esscher Transforms," *Transactions of the Society of Actuaries* 46: 99-191.

Gerber, Hans U. and Shiu, Elias, S.W. 1996. "Actuarial Bridges to Dynamic Hedging and Option Pricing," *Insurance: Mathematics and Economics* 18: 183-218.

Hallman, Victor, G. and Hamilton, Karen, L. 1994. *Personal Insurance: Life, Health, and Retirement.* 1st ed. American Institute for CPCU. Malvern, PA.

Hull, John, C. 1993. *Options, Futures, and Other Derivative Securities.* 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Kellison, Stephen, G. 1991. *Theory of Interest.* Richard D. Irwin, Inc. Homewood, IL.

Lintner, J. 1965. "The Valuation of Risk Assets and the Selection of Risky Investment in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics,* 47: 13-37.

Luenberger, David, G. 1998. *Investment Science.* New York, NY: Oxford University Press.

Markowitz, Harry, M. 1952. "Portfolio Selection." *Journal of Finance*: 77-91.

Markowitz, Harry, M. 1959. *Portfolio Selection: Efficient Diversification of Investments.* New York, NY: John Wiley & Sons.

McGill, Dan, M. et al. eds. 1996. *Fundamentals of Private Pensions.* 7th ed. Philadelphia, PH: University of Pennsylvania Press.

Mitchell, Olivia, S. and Hustead, Edwin, C. 2001. *Pensions in the Public Sector.* Philadelphia: University of Pennsylvania Press.

Mossin, J. 1966. "Equilibrium in a Capital Asset Market," *Econometrica*, 34, no. 4: 768-783.

Panjer, Harry, H. et al eds. 1998. *Financial Economics: With Applications to Investments, Insurance and Pensions.* Schaumburg, IL: The Actuarial Foundation.

Rejda, George, E. 1994. *Social Insurance and Economic Security.* New Jersey: Prentice Hall.

Ross, Sheldon, M. 2000. *Introduction to Probability Models.* 7th ed. San Diego, CA: Academic Press.

Scahill, Patricia, L. 2002. "Introduction to Retirement Income Security Systems." *Course 5 Study Note.* Schaumburg, IL: Society of Actuaries.

Sharpe, W. F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19: 425-442.

World Bank 1994. *Averting the Old Age Crisis.* New York, NY: Oxford University Press.