An Overview of Asset Pricing Models

Andreas Krause

University of Fribourg

Phone: +41-26-3008237
Fax: +41-26-3009700

E-Mail: andreas.krause@unifr.ch

Version of January 6, 2001

Contents

List of Figures

List of Tables

This book gives an overview of the most widely used theories in asset pricing and some more recent developments. The aim of these theories is to determine the *fundamental value* of an asset. As we will see in the first section there is a close relation between this fundamental value and an *appropriate return*. The main focus of asset pricing theories, and therefore of most sections in this chapter, is to determine this appropriate return. The last sections will also show how deviations from the fundamental value can be explained. As the main focus of this chapter is on the theories, empirical investigations are only presented in very short, citing the results of the most prominent works.

The fundamental value of an asset has to be distinguished from its price that we can observe in the market. The *fundamental value* can be identified with the *natural price* as defined by Adam Smith. He defines the natural price to be such that it gives the owner a sufficient profit.[1]

The *price* that can be observed can be interpreted as the *market price* in the sense of Adam Smith. The market price is determined by demand and supply of the asset and can therefore deviate from the fundamental value, but in the long run will converge to the fundamental value.[2]

Although the focus of most theories is laid on the fundamental value asset pricing theories are widely used to explain observed prices. As several theories failed to explain prices sufficiently well they were modified to fit better with the data. This gave rise to a shift in recent years from determinating the fundamental value to explaining prices. With the emergence of the efficient markets hypothesis a close relation between the fundamental value and the price has been proposed, suggesting that the price should always equal the fundamental value. Therefore

[1] See SMITH (1776, p.72). Adam Smith used this definition to characterize the natural price of commodities. Therefore he had also to take into account labor costs and land rents. When applying this definition to assets, we only have to consider profits or equivalently returns. He also mentions that sufficient profits depend on several characteristics of the commodity. It will turn out that the most important characteristic is the risk of an asset.

[2] See SMITH (1776, pp.73 ff.). Empirically there is strong evidence that asset prices (and returns) deviate from the fundamental value (and appropriate return) substantially in the short run, but in the long run follow the fundamental value.

nowadays it is in many cases not distinguished between the determination of the fundamental value and explaining observed prices. In many cases they are looked alike.

In the sections that give a short overview of empirical results concerning the different models it therefore has to be kept in mind that the aim of most models is not to explain prices. The results have to be interpreted as how well the fundamental value of the assets explains the observed prices and not how well the model explains prices. Only the last two models on speculation and financial disequilibria want to explain prices rather than the fundamental value.

1. The Present Value Model

In chapter 2 an asset has been defined as a right on future cash flows. It therefore is straightforward to assume that the value of an asset depends on these cash flows. As we concentrate our analysis on the cash flow of an investor, it consists of the dividends paid by the company, neglecting capital repayments and other infrequent forms of cash flows received by investors.[1]

It is assumed that dividends are paid at the beginning of a period, while the asset can only be bought and sold at the end of a period.[2] This convention can be justified by recalling that a dividend is paid from the company's earnings in the previous period (in most cases a quarter or year) and therefore should be assigned to the holder of the asset in this period.

The following definition links dividends and prices:[3]

$$R_{t+1} \equiv \frac{P_{t+1} - P_t}{P_t} + \frac{D_{t+1}}{P_t}, \tag{1.1}$$

where R_{t+1} denotes the rate of return of the asset from period t to period $t+1$, P_t the price of this asset in period t and D_{t+1} the dividend paid at the beginning of period $t+1$. The first term on the right side represents the capital gain and the

[1] If we assume the company to retain only the part of their cash flow that it can invest as efficient as their shareholders it is of no relevance to the shareholders whether a dividend is paid or not. If cash flow is retained the value of the asset increases. This increase in the value of the company compensates the investor for receiving only a part of the company's cash flow in form of a dividend. It is therefore reasonable not to distinguish between the cash flow of a company and the dividend received by the investor. If the company invests the cash flow less or more efficient than an investor would do, the value of the company is reduced in the first and increased in the second case. Throughout this chapter we neglect any behavior of the companies or the management that can harm the investors, i.e. we assume that no agency problem exists.

[2] See Campbell et al. (1997, pp.254f.).

[3] See Campbell et al. (1997, p.12).

second term the dividend yield. Solving this definition for P_t gives a difference equation for the price in period t:

$$P_t = \frac{P_{t+1} + D_{t+1}}{1 + R_{t+1}}. \tag{1.2}$$

Solving this difference equation forward for k periods results in

$$P_t = \sum_{i=1}^{k} \left[\prod_{j=1}^{i} \frac{1}{1 + R_{t+j}} \right] D_{t+i} + \left[\prod_{j=1}^{k} \frac{1}{1 + R_{t+j}} \right] P_{t+k}. \tag{1.3}$$

If we assume the asset price to grow at a lower rate than R_{t+j}, the last term converges to zero:[4]

$$\lim_{k \to \infty} \left[\prod_{j=1}^{i} \frac{1}{1 + R_{t+j}} \right] P_{t+k} = 0. \tag{1.4}$$

Using (1.4) to increase k in (1.3) we obtain the price to be the *present value* of the dividends:

$$P_t = \sum_{i=1}^{\infty} \left[\prod_{j=1}^{i} \frac{1}{1 + R_{t+j}} \right] D_{t+i} \tag{1.5}$$

Equation (1.5) only holds *ex-post*, i.e. all future dividends and rates of return have to be known for the determination of the current price. In the following section it will be shown how this formula can be used to determine the price without knowing future dividends and rates of return.

The following result from Campbell et al. (1997, p.253) can easily be verified:

The price of the asset changes more with a persistent movement in dividends or rates of return than a temporary movement of the same amount. An increase in dividends and a decrease in the rate of return increases the current price.

1.1 The Efficient Market Hypothesis

The definition of an efficient market is given by Fama (1970, p.383):

[4] See Campbell et al. (1997, p.255) and Aschinger (1995, pp. 39f.). A solution for the case that this condition is not fulfilled is provided in section 4.11.

> "A market in which prices always "fully reflect" available information is called "efficient"."

With this definition in an efficient market the price should always equal the fundamental value that is determined according to the information available. Sufficient, but not necessary, conditions for a market to be efficient are:[5]

- no transaction costs for trading the asset,
- all information is available at no costs for all market participants,
- all market participants agree in the implications information has on current and future prices and dividends.

Three forms of efficiency are distinguished in the literature: *weak*, *semistrong* and *strong* efficiency. These forms differ only in the set of information that has to be incorporated into prices. Weak efficiency uses only information on past prices and returns, semistrong efficiency includes all public available information and the strong form includes all information available to any market participant including private information.[6]

Let the information set available in period t be denoted Ω_t. In the further discussion the different forms of efficiency are not distinguished, Ω_t represents the information set that suites the form needed.

Given this information set, a price *fully reflects* information if, based on this information, no market participant can generate expected profits higher than the equilibrium profit:[7]

$$E\left[x_{t+1}|\Omega_t\right] = 0, \qquad (1.6)$$

[5] See Fama (1970, p.387).

[6] See Fama (1970, p.383), Campbell et al. (1997, p.22) and Aschinger (1995, p.40).

[7] See Fama (1970, p. 385). This also implicitly assumes that all market participants make use of their information to form expectations (rational expectations).

where

$$x_{t+1} = R_{t+1} - E\left[R_{t+1}|\Omega_t\right].$$

Given the information set the expected return $E\left[R_{t+1}|\Omega_t\right]$ equals the realized return R_{t+1} on average, i.e. there are no systematic errors in predicting future returns that could be used to make extraordinary profits. Using the (1.2) we see that

$$P_{t+1} + D_{t+1} = \left(1 + R_{t+1}\right) P_t. \tag{1.7}$$

Taking conditional expectations on both sides and rearranging the terms we get[8]

$$P_t = \frac{E\left[P_{t+1}|\Omega_t\right] + E\left[D_{t+1}|\Omega_t\right]}{1 + E\left[R_{t+1}|\Omega_t\right]}. \tag{1.8}$$

The expected return $E\left[R_{t+1}|\Omega_t\right]$ is a function of the risk of the asset and has to be determined by a separate theory.[9]

When comparing equations (1.2) and (1.8) we see that they differ only by using expected values instead of the realized values for future prices, dividends and rates of return. With the assumption of efficient markets we see from (1.6) that an investor can make no extraordinary profits if the asset is priced according to (1.8).[10] Hence we are able to derive an appropriate price *ex ante* by forming expectations.

The same manipulations and assumptions that were made for the derivation of equation (1.5) gives us the expression for the fundamental value of the asset:

$$P_t = \sum_{i=1}^{\infty} \left[\prod_{j=1}^{i} \frac{1}{1 + E\left[R_{t+j}|\Omega_t\right]}\right] E\left[D_{t+i}|\Omega_t\right]. \tag{1.9}$$

[8] It is necessary to transform (1.2) into (1.7) first to obtain a linear form. This linear form allows to take expectations and receive the desired result.

[9] See Fama (1970, p.384). Several of these theories are presented in subsequent sections.

[10] Of course, the assumptions of the efficient markets hypothesis have to be fulfilled. Especially if the investors differ in their beliefs of the implications information has or if they have different information, individual investors may be able to make expected extraordinary profits.

The fundamental value depends on expected future dividends and rates of return. The expected dividends are discounted with the expected rates of returns, i.e. the present values of all expected future dividends are added up. That is why this model is called the *present value model*.

While for the determination of the expected returns theories exist that will be presented in the following sections, for dividends no such theories exist. Which dividends are paid by the companies depends on the success of the company and has to be analyzed individually for each company. We will therefore impose some restrictions on the expectations to get further insights into the properties of the model.

1.2 The random walk model

A more restrictive version of the efficient market hypothesis is the *random walk model*, which assumes successive price changes to be independent of each other. Additionally it is assumed that these successive price changes are identically distributed.[11] In this case we have for all $i = 1, 2, \ldots$:

$$E\left[R_{t+i}|\Omega_t\right] = E\left[R_{t+i}\right] = R_t, \tag{1.10}$$

i.e. conditional and unconditional expected values of the return are identical on constant over time.[12]

Inserting (1.10) into (1.9) we get

$$P_t = \sum_{i=1}^{\infty}\left(\frac{1}{1+R_t}\right)^i E\left[D_{t+i}|\Omega_t\right]. \tag{1.11}$$

Two special cases concerning the behavior of expected dividends can frequently be found in the literature: constant expected dividends and constant growth rates of expected dividends.

[11] See Fama (1970, p.386).

[12] This assumption is the basis for most asset pricing theories, only more recent models allow returns to vary over time.

With constant expected dividends, i.e. $E[D_{t+i}|\Omega_t] = D_t$ for all $i = 1, 2, \ldots$ equation (1.11) reduces to

$$P_t = \frac{D_t}{R_t}, \tag{1.12}$$

the fundamental value equals the capitalized dividend. This formula can frequently be found in corporate finance to determine the value of a company or an investment.[13] The dependence of the fundamental value on the dividend and an appropriate return can easily be seen from this formula: the higher the dividend and the lower the appropriate return the higher the price of the asset is.

Another restrictive model is the *Gordon growth model*.[14] This model assumes the expected dividends to grow with a fixed rate of $G_t < R_t$:

$$E[D_{t+i+1}|\Omega_t] = (1 + G_t)E[D_{t+i}|\Omega_t] = (1 + G_t)^{i+1} D_t. \tag{1.13}$$

Inserting this relation into (1.11) we obtain

$$P_t = \frac{1 + G_t}{R_t - G_t} D_t. \tag{1.14}$$

This relation gives the insight that the price does not only depend on the current dividend, but also on the growth of the dividend. Especially if the growth rate is near the expected return the price is very sensitive to a change in the growth rate:[15]

$$\frac{\partial P_t}{\partial G_t} = \frac{1 + R_t}{(R_t - G_t)^2} D_t.$$

This formula, although the assumptions are very restrictive, gives a justification of relative high prices for tech stocks. Many of these companies pay only a small dividend at present (if at all), but they are expected to grow very fast. This growth rate justifies the high price-earnings ratio observed in markets.[16]

[13] See BOEMLE (1993, pp. 604ff.).
[14] See GORDON (1962, pp. 43ff.).
[15] A growth rate above expected returns would give an infinite price. This case is similar to a violation of equation (1.4) and is therefore excluded here.
[16] Consider the following example: A company pays a current dividend of $D_t = .1\$$ per share, the expected return is $R_t = .25$ and the growth rate is $G_t = .2$. According to (1.12) the price should be $P_t = .4\$$. But taking into account the growth rate, (1.14) suggests a price

A more general form of the Gordon model can be obtained by manipulating (1.11) further.[17] Shifting the formula one period ahead and taking conditional expectations gives

$$\begin{aligned} E\left[P_{t+1}|\Omega_t\right] &= E\left[\sum_{i=1}^{\infty}\left(\frac{1}{1+R_t}\right)^i E\left[D_{t+i+1}|\Omega_t\right]\middle|\,\Omega_t\right] \\ &= \sum_{i=1}^{\infty}\left(\frac{1}{1+R_t}\right)^i E\left[D_{t+i+1}|\Omega_t\right]. \end{aligned}$$

Subtracting

$$P_t + D_t = \sum_{i=0}^{\infty}\left(\frac{1}{1+R_t}\right)^i E\left[D_{t+i}|\Omega_t\right]$$

results in

$$E\left[P_{t+1}|\Omega_t\right] - P_t - D_t = \sum_{i=1}^{\infty}\left(\frac{1}{1+R_t}\right)^i E\left[D_{t+i+1} - D_{t+i}|\Omega_t\right] - D_t.$$

By substituting $E\left[P_{t+1}|\Omega_t\right] = (1+R_t)P_t + E\left[D_{t+1}|\Omega_t\right]$ from (1.7) we get

$$\begin{aligned} R_tP_t - D_t &= \sum_{i=1}^{\infty}\left(\frac{1}{1+R_t}\right)^i E\left[D_{t+i+1} - D_{t+i}|\Omega_t\right] + E\left[D_{t+1} - D_t|\Omega_t\right] \\ &= \sum_{i=0}^{\infty}\left(\frac{1}{1+R_t}\right)^i E\left[D_{t+i+1} - D_{t+i}|\Omega_t\right] \\ &= \sum_{i=0}^{\infty}\left(\frac{1}{1+R_t}\right)^i E\left[\Delta D_{t+i+1}|\Omega_t\right], \end{aligned}$$

where $\Delta D_{t+1} = D_{t+1} - D_t$ denotes the change in the dividend from period t to period $t+1$. Rearranging we receive

$$P_t = \frac{D_t}{R_t} + \frac{1}{R_t}\sum_{i=0}^{\infty}\left(\frac{1}{1+R_t}\right)^i E\left[\Delta D_{t+i+1}|\Omega_t\right]. \tag{1.15}$$

The fundamental value consists according to this representation of the capitalized value of the current dividend and the capitalization of the present value of all

of $P_t = 2.4\$$. The dividend yield is only 4% and a price earnings ratio of 24 if we assume all earnings to be paid as dividends to the investors. Another company working in a much less growing, but also less risky sector also pays a dividend of $D_t = .1\$$ per share, with an expected return of $R_t = .15$ the value according to (1.12) is $P_t = .67\$$ with a growth rate of $G_t = .05$, while for the Gordon growth model it is $P_t = 1.05\$$, giving a price earnings ratio of only 10.5. Although the expected return of the first company is much higher, it has a higher value due to its fast growth.

[17] See Campbell et al. (1997, p.257).

subsequent expected changes in the dividend. Equation (1.15) allows more general assumptions on the dividend growth process than does the Gordon growth model.

Not only the level of dividends, but also the expected growth has a significant influence on the fundamental value. Further insights can be expected from returning to the general case where expected returns are allowed to change over time.

1.3 The dynamic Gordon growth model

Evidence from more recent asset pricing models (see e.g. chapter 4.4) suggests that expected returns are not constant over time. Along with the observation that dividends are also varying over time this gave rise to the *dynamic Gordon growth model* as presented by Campbell/Shiller (1988b) and Campbell/Shiller (1988a).

The non-linearity in expected returns of equation (1.9) makes this formula difficult to evaluate. We therefore approximate this relation by a log-linear model. Let us denote

$$\begin{aligned} r_{t+1} &= \ln\left(1 + E\left[R_{t+1}|\Omega_t\right]\right), \\ p_{t+i} &= \ln E\left[P_{t+i}|\Omega_t\right], \\ d_{t+i} &= \ln E\left[D_{t+i}|\Omega_t\right]. \end{aligned} \tag{1.16}$$

From (1.8) we get after taking logs

$$\begin{aligned} p_t &= \ln E\left[P_{t+1} + D_{t+1}|\Omega_t\right] - r_{t+1} \\ &= \ln E\left[P_{t+1}\left(1 + \frac{D_{t+1}}{P_{t+1}}\right)\middle|\,\Omega_t\right] - r_{t+1} \\ &= p_{t+1} - r_{t+1} + \ln\left(1 + \exp\left(d_{t+1} - p_{t+1}\right)\right). \end{aligned} \tag{1.17}$$

The last term can be approximated by a first order Taylor series expansion around

$E[d_{t+1} - p_{t+1}]$:

$$\ln(1+\exp(d_{t+1}-p_{t+1})) \approx \ln(1+\exp(E[d_{t+1}-p_{t+1}])) + \frac{\exp(E[d_{t+1}-p_{t+1}])}{1+\exp(E[d_{t+1}-p_{t+1}])}(d_{t+1}-p_{t+1}-E[d_{t+1}-p_{t+1}]). \tag{1.18}$$

Inserting this expression and rearranging gives

$$\begin{aligned} p_t &= p_{t+1}\left(1-\frac{\exp(E[d_{t+1}-p_{t+1}])}{1+\exp(E[d_{t+1}-p_{t+1}])}\right) \\ &\quad + \frac{\exp(E[d_{t+1}-p_{t+1}])}{1+\exp(E[d_{t+1}-p_{t+1}])}d_{t+1} \\ &\quad + \ln(1+\exp(E[d_{t+1}-p_{t+1}])) \\ &\quad - \frac{E[d_{t+1}-p_{t+1}]\exp(E[d_{t+1}-p_{t+1}])}{1+\exp(E[d_{t+1}-p_{t+1}])} - r_{t+1}. \end{aligned} \tag{1.19}$$

By defining

$$\begin{aligned} \rho &= \frac{1}{1+\exp(E[d_{t+1}-p_{t+1}])}, \\ k &= (1-\rho)\ln(\rho-1) - \ln\rho \end{aligned}$$

this gives

$$\begin{aligned} p_t &= p_{t+1}\rho + (1-\rho)d_{t+1} + \ln\left(\frac{1}{\rho}\right) - (1-\rho)\ln\left(\frac{1}{\rho-1}\right) - r_{t+1} \\ &= \rho p_{t+1} + (1-\rho)d_{t+1} + (1-\rho)\ln(\rho-1) - \ln\rho - r_{t+1} \\ &= k + \rho p_{t+1} + (1-\rho)d_{t+1} - r_{t+1}. \end{aligned} \tag{1.20}$$

It is now assumed that ρ is constant over time, i.e. the expected dividend-price ratio does not vary over time. Defining δ_t as the expected log dividend price ratio and Δd_t as the expected change in log dividends:

$$\delta_t = d_t - p_t, \tag{1.21}$$

$$\Delta d_t = d_{t+1} - d_t \tag{1.22}$$

we get from the above relation

$$\begin{aligned} r_{t+1} &= k - p_t + \rho p_{t+1} + (1-\rho)d_{t+1} \\ &= k + d_t - p_t - \rho(d_{t+1} - p_{t+1}) + d_{t+1} - d_t \\ &= k + \delta_t - \rho\delta_{t+1} + \Delta d_{t+1}. \end{aligned} \tag{1.23}$$

If we in analogy to (1.5) assume that

$$\lim_{i \to \infty} \rho^i \delta_{t+i} = 0 \tag{1.24}$$

we can solve differential equation (1.23) forward and obtain

$$\delta_t = \sum_{j=0}^{\infty} \rho^j \left(r_{t+j} - \Delta d_{t+j} \right) - \frac{k}{1-\rho}. \tag{1.25}$$

Substituting back into (1.21) gives the pricing formula of the dynamic Gordon growth model:

$$p_t = d_t + \frac{k}{1-\rho} + \sum_{j=0}^{\infty} \rho^j \left(r_{t+j} - \Delta d_{t+j} \right). \tag{1.26}$$

The log price has been found to be a linear combination of the present value of the increase in expected dividends, future expected returns, the current dividend and a constant factor that adopts for the level of prices.[18]

The price depends not only on the growth rate of the dividends, but also on how long this growth is expected to be effective and how long the expected returns change.[19] This result has already been derived in the more general approach without approximating by a log-linear function.

The advantage of this approximation compared to the exact formula (1.9) is its linear structure. This makes it much more easy to use in an empirical investigation. The next section provides a short overview of the results obtained from such investigations.

1.4 Empirical results

An empirical investigation will face two major problems as can be seen by inspection of equation (1.26). The first problem is the infinite sum. In applying

[18] See CAMPBELL/SHILLER (1988b, p.199).
[19] See CAMPBELL ET AL. (1997, p.263).

the formula we have to terminate the sum after $T < \infty$ terms, but with assumption (1.24) this effect should not violate our results too much as for large T the remaining terms are close to zero.

The second difficulty that arises is that only realized variables can be observed, whereas we need the expected values of these variables. With the assumption of an efficient market it is reasonable to assume that on average the expectations are fulfilled, i.e. remaining errors should be small if the investigation includes enough time periods.

Without going into details how to conduct an empirical investigation,[20] it turns out that dividends are a good forecast for future prices (or equivalently returns) for a time horizon of at least four years. For shorter time horizons the explanatory power of dividends decreases fast, for one year dividends can only explain about 1 percent of the returns.[21]

Campbell/Shiller (1988a) show that forecasted prices using the present value model are much smoother than realized prices, but the pattern of the price changes seems to follow the same scheme.[22]

The results show, together with many other empirical investigations, that in the long term the prices behave approximately as predicted by the present-value model. But in the short term prices can deviate substantially from the model.

The next sections will therefore investigate more carefully the expected returns of an asset to explain these short term deviations from the present value.

[20] See Campbell et al. (1997) for an overview.
[21] See Campbell et al. (1997, p.269).
[22] See also Campbell et al. (1997, p.283).

2. The Portfolio Selection Theory

When considering to invest into the stock market,[1] an investor has to make three decisions:

- the amount he wants to invest into the stock market,
- determine the stocks he wants to invest in,
- the amount he wants to invest into each selected stock.

This section describes a method how to make these decisions and find an *optimal portfolio*.[2] Such a portfolio

> " ... is more than a long list of good stocks and bonds. It is a balanced whole, providing the investor with protections and opportunities with respect to a wide range of contengencies. The investor should build toward an integrated portfolio which best suites his needs."[3]

For this reason the associated theory is called *portfolio selection theory* or short *portfolio theory*, rather than asset selection.[4] The portfolio selection theory has been developed by Markowitz (1959), Tobin (1958) and Tobin (1966). Although the concepts employed in their theory have much been criticized for capturing the reality only poorly, it has been the starting point for many asset pricing models and up to date there has been developed no widely accepted alternative.

[1] Other assets, such as real estate, commodities or human capital could easily be included. As we concentrate on stock markets in this thesis, we only consider stocks here.
[2] A *portfolio* is the entirety of all investments of an individual. See Büschgen (1991, p.552).
[3] Markowitz (1959, p.3).
[4] See Markowitz (1959, p.3).

Before describing the theory in detail we will at first work out the utility concept that is used in the portfolio selection theory and throughout most literature on financial markets.

2.1 The utility concept applied in portfolio theory

As we have seen in the last section the value, and therewith the returns of assets depend on their future dividends. These future dividends cannot be predicted with certainty, it is a random variable, hence the return is also a random variable from the view of an investor. The decisions mentioned above have to be made under risk.[5] In their work von Neumann/Morgenstern (1953) presented a criterion to make an optimal decision if five axioms are fulfilled.[6]

Let $A = \{a_1, \ldots, a_N\}$ be the set of all possible alternatives an individual can choose between,[7] $S = \{s_1, \ldots, s_M\}$ all states that affect the outcome,[8] and $C = \{c_{11}, \ldots, c_{1M}; \ldots; c_{N1}, \ldots, c_{NM}\}$ the outcomes, where c_{ij} is the outcome if state s_j occurs and alternative a_i has been chosen.

Axiom 1. *A is completely ordered.*

A set is completely ordered if it is complete, i.e. we have either $a_i \succeq a_j$ or $a_j \succeq a_i$ for all $i, j = 1, \ldots, N$ where "$\succeq$" denotes the preference. The set has further to be transitive, i.e. if $a_i \succeq a_j$ and $a_j \succeq a_k$ we then have $a_i \succeq a_k$. Axiom 1 ensures that all alternatives can be compared with each other and are ordered

[5] According to Knight (1921, pp. 197ff.) a decision has to be made under *risk* if the outcome is not known with certainty, but the possible outcomes and the probabilities of each outcome are known. The probabilities can either be assigned by objective or subjective functions. Keynes (1997, p. 68) defines risk as the possibility of the actual outcome to be different from the expected outcome.

[6] Many different ways to present these axioms can be found in the literature. We here follow the version of Levy/Sarnat (1972, p. 202)

[7] As for $N = 1$ there is no decision to make for the individual it is required that $N \geq 2$.

[8] As with $M = 1$ the outcome can be predicted with certainty we need $M \geq 2$.

consistently.[9]

To state the remaining axioms we have to introduce some notations. Let any alternative be denoted as a lottery, where each outcome c_{ij} has a probability of p_{ij}. We can write alternative i as

$$a_i = [p_{i1}c_{i1}, \ldots, p_{iM}c_{iM}] \text{ , where } \sum_{j=1}^{M} p_{ij} = 1 \text{ for all } i = 1, \ldots, N.$$

Axiom 2 (Decomposition of compound lotteries). *If the outcome of a lottery is itself a lottery (compound lottery), the first lottery can be decomposed into its final outcomes:*

Let $a_i = [p_{i1}b_{i1}, \ldots, p_{iM}b_{iM}]$ *with* $b_{ij} = [q_{ij1}c_1, \ldots, q_{ijL}c_L]$. *With* $p^*_{ik} = \sum_{l=1}^{M} p_{il}q_{ilk}$[10] *we have* $[p_{i1}b_{i1}, \ldots, p_{iM}b_{iM}] \sim [p^*_{i1}c_1, \ldots, p^*_{iM}c_L]$

Axiom 3 (Composition of compound lotteries). *If an individual is indifferent between two lotteries they can be interchanged into a compound lottery:*

If $a_i = [p_{i1}b_{i1}, \ldots, p_{ij}b_{ij}, \ldots, p_{iM}b_{iM}]$ *and* $b_{ij} \sim [q_{ij1}c_1, \ldots, p_{ijL}c_L]$ *then* $a_i \sim [p_{i1}b_{i1}, \ldots, p_{ij}[q_{ij1}c_1, \ldots, p_{ijL}c_L], \ldots, p_{iM}b_{iM}]$.

These two axioms ensure that lotteries can be decomposed into their most basic elements (axiom 4.2) and that more complex lotteries can be build up from their basic elements (axiom 4.3).

Axiom 4 (Monotonicity). *If two lotteries have the same two possible outcomes, then the lottery is preferred that has the higher probability on the more preferred outcome:*

Let $a_i = [p_{i1}c_1, p_{i2}c_2]$ *and* $b_i = [q_{i1}c_1, q_{i2}c_2]$ *with* $c_1 \succ c_2$, *if* $p_{i1} > q_{i1}$ *then* $a_i \succ b_i$.

[9] The transitivity ensures consistent decisions of individuals. It is equivalent with the usual assumption in microeconomics that indifference curves do not cross. See SCHUMANN (1992, pp. 52 ff.).

[10] This representation of joint probabilities assumes that the two lotteries are independent. If the two lotteries where not independent the formula has to be changed, but the results remain valid. It is also assumed throughout this chapter that there is no joy of gambling, i.e. that there is no gain in utility from being exposed to uncertainty.

Given the same possible outcomes this axiom ensures the preference relation"$\succ$" to be a monotone transformation of the relation "$>$" between probabilities.

Axiom 5 (Continuity). *Let a_i, b_i and c_i be lotteries. If $a_i \succ b_i$ and $b_i \succ c_i$ then there exists a lottery d_i such that $d_i = [p_1 a_i, p_2 c_i] \sim b_i$.*

This axiom ensures the mapping from the preference relation "$\succ$" to the probability relation "$>$" to be continuous.

The validity of these axioms is widely accepted in the literature. Other axioms have been proposed, but the results derived from these axioms are identical to those to be derived in an instant.[11]

Given these assumptions the following theorem can be derived, where U denotes the utility function.

Theorem 1 (Expected utility principle). *An alternative a_i will be preferred to an alternative a_j if and only if the expected utility of the former is larger, i.e*

$$a_i \succ a_j \Leftrightarrow E\left[U\left(a_i\right)\right] > E\left[U\left(a_j\right)\right].$$

Proof. Define a lottery $a_i = [p_{i1}c_1, \ldots, p_{iM}c_M]$, where without loss of generality $c_1 \succ c_2 \succ \cdots \succ c_M$. Such an order is ensured to exist by axiom 4.1.

Using axiom 4.5 we know that there exists a lottery such that

$$c_i \sim [u_{i1}c_1, u_{i2}c_M] = [u_i c_1, (1-u_i)c_M] \equiv c_i^*.$$

We can now use axiom 4.3 to substitute c_i by c_i^* in a_i:

$$a_i \sim [p_{i1}c_1^*, \ldots, p_{iM}c_M^*].$$

This alternative only has two possible outcomes: c_1 and c_M. By applying axiom 4.2 we get

[11] See Markowitz (1959, p. xi).

$a_i \sim [p_i c_1, (1-p_i)c_M]$ with $p_i = \sum_{j=1}^{M} u_i p_{ij}$,

what is the definition of the expected value for discrete random variables: $p_i = E[u_i]$.[12]

The same manipulations as before can be made for another alternative a_j, resulting in

$a_j \sim [p_j c_1, (1-p_j)c_M]$ with $p_j = E[u_j]$.

If $a_i \succ a_j$ then we find with axiom 4.4 that, as $c_1 \succ c_M$:

$$p_i > p_j.$$

The numbers u_{ij} we call the utility of alternative a_i if state s_j occurs. The interpretation as utility can be justified as follows: If $c_i \succ c_j$ then axiom 4.4 implies that $u_i > u_j$, we can use u_i to index the preference of the outcome, i.e. a higher u implies preference for this alternative and vice versa.

Therewith we have shown that $a_i \succ a_j$ is equivalent to $E\left[U\left(a_i\right)\right] > E\left[U\left(a_j\right)\right]$. $\square$

The criterion to choose between two alternatives, in our case which portfolio to choose, is to take that alternative with the highest expected utility. To apply this criterion the utility function has to be known. As in most cases we do not know the utility function, it is necessary to analyze this criterion further to derive a more handable criterion.

2.2 Risk aversion

> "Individuals are *risk averse* if they always prefer to receive a fixed payment to a random payment of equal expected value."[13]

[12] The extension to continuous random variables is straightforward by replacing the probabilities with densities.

[13] Dumas/Allaz (1996, p. 30), emphasize added.

From many empirical investigations it is known that individuals are risk averse, where the degree of risk aversion differs widely between individuals.[14] The Arrow-Pratt measure is the most widely used concept to measure risk aversion. We will derive this measure following Pratt (1964).

With the definition of risk aversion above, an individual prefers to receive a fixed payment of $E[\widetilde{x}]$ to a random payment of $\widetilde{x}$. To make the individual indifferent between a fixed payment and a random payment, there exists a number π, called *risk premium*, such that he is indifferent between receiving $E[\widetilde{x}] - \pi$ and $\widetilde{x}$. By applying theorem 4.6 we see that the expected utility of these two payments has to be equal:

$$E\left[U\left(\widetilde{x}\right)\right] = E\left[U\left(E\left[\widetilde{x}\right] - \pi\right)\right] = U\left(E\left[\widetilde{x}\right] - \pi\right). \tag{2.1}$$

The term $E\left[\widetilde{x}\right] - \pi$ is also called the *cash equivalent* of $\widetilde{x}$. Approximating the left side by a second order Taylor series expansion around $E\left[\widetilde{x}\right]$ we get[15]

$$\begin{aligned} E\left[U\left(\widetilde{x}\right)\right] &= E\Big[U\left(E\left[\widetilde{x}\right]\right) + U'(E[\widetilde{x}])(\widetilde{x} - E[\widetilde{x}]) \\ &\quad + \frac{1}{2}U''(E[\widetilde{x}](\widetilde{x} - E[\widetilde{x}])^2\Big] \\ &= U(E[\widetilde{x}]) + U'(E[\widetilde{x}])E[\widetilde{x} - E[\widetilde{x}]] \\ &\quad + \frac{1}{2}U''(E[\widetilde{x}])E[(\widetilde{x} - E[\widetilde{x}])^2] \\ &= U(E[\widetilde{x}]) + \frac{1}{2}U''(E[\widetilde{x}])Var[\widetilde{x}]. \end{aligned} \tag{2.2}$$

where $U^{(n)}(E[\widetilde{x}])$ denotes the nth derivative of U with respect to its argument evaluated at $E[\widetilde{x}]$. In a similar way we can approximate the right side by a first order Taylor series around $E[\widetilde{x}]$ and get

$$U(E[\widetilde{x}] - \pi) = U(E[\widetilde{x}]) + U'(E[\widetilde{x}])\pi. \tag{2.3}$$

[14] Despite this clear evidence for risk aversion many economic theories assume that the individuals are risk neutral. Prominent examples are the information-based models of market making (see section 5.5) and the Black-Scholes formula for the pricing of options.

[15] We assume higher order terms to be negligable, what can be justified if $\widetilde{x}$ does not vary too much from $E\left[\widetilde{x}\right]$.

Inserting (2.2) and (2.3) into (2.1) and solving for the risk premium π we get

$$\pi = \frac{1}{2}\left(-\frac{U''(E[\widetilde{x}])}{U'(E[\widetilde{x}])}\right)Var(\widetilde{x}). \tag{2.4}$$

Pratt (1964) now defines

$$z = -\frac{U''(E[\widetilde{x}])}{U'(E[\widetilde{x}])} \tag{2.5}$$

as the *absolute local risk aversion*. This can be justified by noting that the risk premia has to be larger the more risk averse an individual is and the higher the risk. The risk is measured by the variance of $\widetilde{x}$, $Var[\widetilde{x}]$,[16] hence the other term in (2.4) can be interpreted as risk aversion. Defining $\sigma^2 = Var[\widetilde{x}]$ we get by inserting (2.5) into (2.4):

$$\pi = \frac{1}{2}z\sigma^2. \tag{2.6}$$

If we assume that individuals are risk averse we need $\pi > 0$, implying $z > 0$. It is reasonable to assume positive marginal utility, i.e. $U'(E[\widetilde{x}]) > 0$, then this implies that $U''(E[\widetilde{x}]) < 0$. This relation is also known as the first Gossen law and states the saturation effect.[17] The assumption of risk aversion is therefore in line with the standard assumptions in microeconomic theory.

The conditions $U'(E[\widetilde{x}]) > 0$ and $U''(E[\widetilde{x}]) < 0$ imply a concave utility function. The concavity of the function (radius) is determined by the risk aversion.[18]

Figure 2.1 visualizes these finding for the simple case of two possible outcomes, x_1 and x_2, having equal probability of occurrence.

[16] For a justification of measuring risk by a variance see the following section
[17] See Schumann (1992, p. 49).
[18] For risk neutral individuals the risk premium, and hence the risk aversion, is zero, resulting in a zero second derivative of U, the utility function has to be linear. For risk loving individuals the risk premium and the risk aversion are negative, hence the second derivative of the utility function has to be positive, hence it is convex. See e.g. Beekmann (1995, p. 6.4/12) for the concept of concavity and convexity.

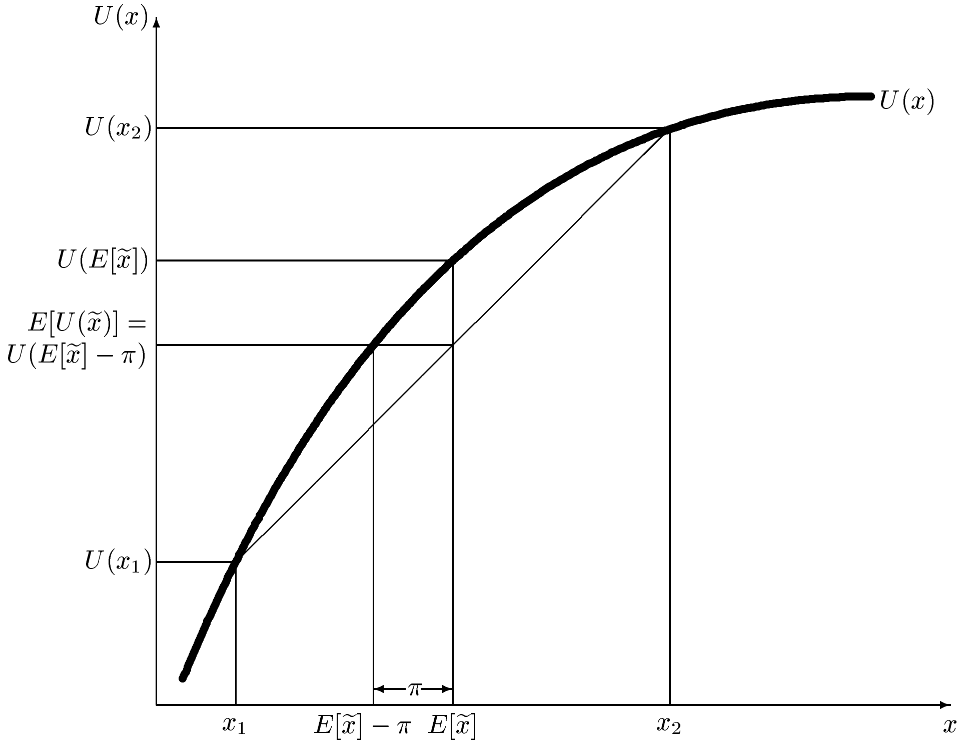

Figure 2.1: The Arrow-Pratt measure of risk aversion

2.3 The Mean-Variance criterion

Even by using the Arrow-Pratt measure of risk aversion, the utility function has to be known to determine the first and second derivative for basing a decision on the expected utility concept. Preferable would be criterion that uses only observable variables instead of individual utility functions. For this purpose many criteria have been proposed, the most widely used is the mean-variance criterion, although it also is not able to determine the optimal decision it restricts the alternatives to choose between by using the utility function.[19]

The *mean-variance criterion* has become popular first as it is not difficult to apply and has some nice properties in terms of moments of a distribution and secondly by the use of this criterion in the basic works on portfolio selection by Markowitz (1959), Tobin (1958), and Tobin (1966). Consequently theories basing on their work, like the Capital Asset Pricing Model, also apply the mean-

[19] See Levy/Sarnat (1972, ch. VII and ch. IX) for an overview of these criteria.

variance criterion, which by these means became the most widely used criterion in finance.

It has the advantage that only two moments of the distribution of outcomes, mean and variance, have to be determined, whereas other criteria make use of the whole distribution.[20] The outcome is characterized by its expected value, the mean, and its risk, measured by the variance of outcomes.[21]

The mean-variance criterion is defined as

$$a_i \gtrsim a_j \Leftrightarrow \begin{cases} Var[a_i] < Var[a_j] \quad \text{and} \quad E[a_i] \geq E[a_j] \\ \text{or} \\ Var[a_i] \leq Var[a_j] \quad \text{and} \quad E[a_i] > E[a_j] \end{cases} . \tag{2.7}$$

It is necessary, although not sufficient, condition to prefer a_i over a_j that $Var[a_i] \leq Var[a_j]$ and $E[a_i] \geq E[a_j]$. An alternative is preferred over another if it has a smaller risk (variance) and a larger mean. Nothing can in general be said about the preferences if $Var[a_i] > Var[a_j]$ and $E[a_i] > E[a_j]$, other decision rules have to be applied.[22]

In figure 2.2 an alternative in the shaded areas can be compared to a_i by using the mean-variance criterion, while in the white areas nothing can be said about the preferences. If we assume all alternatives to lie in a compact and convex set in the (μ, σ^2)-plane,[23] all alternatives that are not dominated by another alternative according to the mean-variance criterion lie on a line at the upper left of the set of alternatives. In figure 2.3 this is illustrated where all alternatives are located in the oval. The undominated alternatives are represented by the bold line between points A and B. All alternatives that are not dominated by another

[20] See Levy/Sarnat (1972, pp. 307 ff.).

[21] One of the main critics of the mean-variance criterion starts with the assumption that risk can be measured by the variance. Many empirical investigations have shown that the variance is not an appropriate measure of risk. Many other risk measures have been proposed, see Brachinger/Weber (1996) for an overview, but these measures have the disadvantage of being less easily computable and difficult to implement as a criterion. In more recent models higher moments, such as skewness and kurtosis are also incorporated to cover the distribution in more detail.

[22] See Levy/Sarnat (1972, pp. 308).

[23] We will see that this condition is fulfilled in the case of portfolio selection for all relevant portfolios.

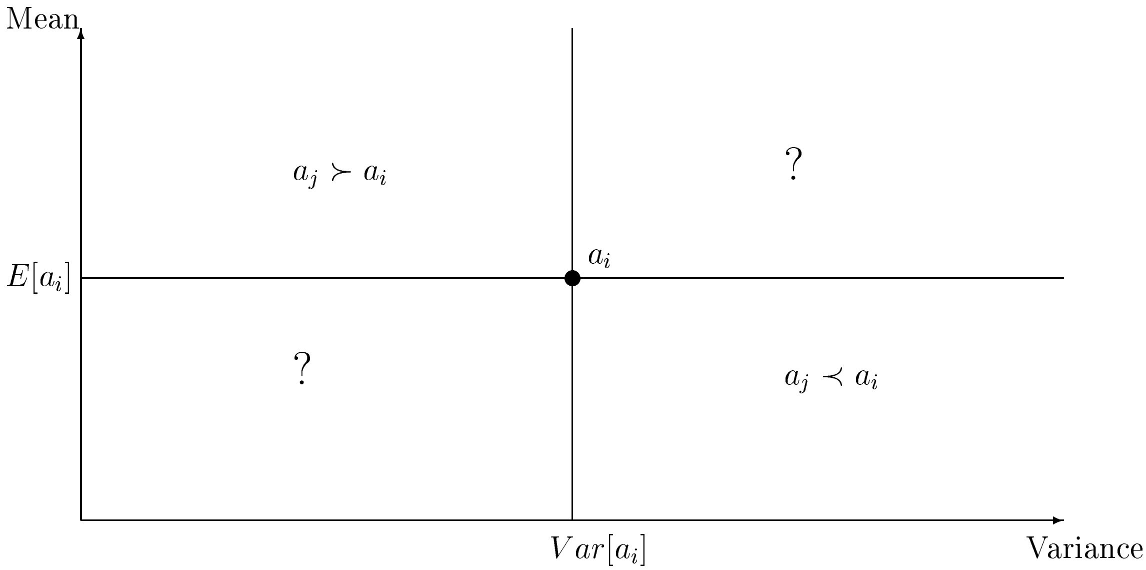

Figure 2.2: The mean-variance criterion

alternative are called *efficient* and all efficient alternatives form the *efficient frontier*.[24] Without having additional information, e.g. the utility function, between efficient alternatives cannot be distinguished.

The mean-variance criterion can be shown to be not optimal in general, i.e. the true preferences are not always reflected by the results of this criterion.[25] If the utility function is quadratic, we will show that the mean-variance criterion always reflects the true preferences.[26]

Instead of defining the utility function by a term like $y = b_0 + b_1 x + b_2 x^2$ we can without loss of generality normalize the function by choosing $b_0 = 0$ and $b_1 = 1$.[27]

[24] See LEVY/SARNAT (1972, pp. 318 ff.).
[25] See LEVY/SARNAT (1972, pp. 310 f.). They also provide a generalization of the mean-variance criterion that is always optimal. As this criterion cannot be handled so easily it is rarely applied and therefore not further considered here.
[26] See LEVY/SARNAT (1972, pp. 379 ff.).
[27] The concept of expected utility implies that the utility function is only determined up to a positive linear transformation. This allows to apply the transformation $y \rightarrow \frac{y-b_0}{b_1}$ and achieve the normalization. See LEVY/SARNAT (1972, pp. 205 and 379).

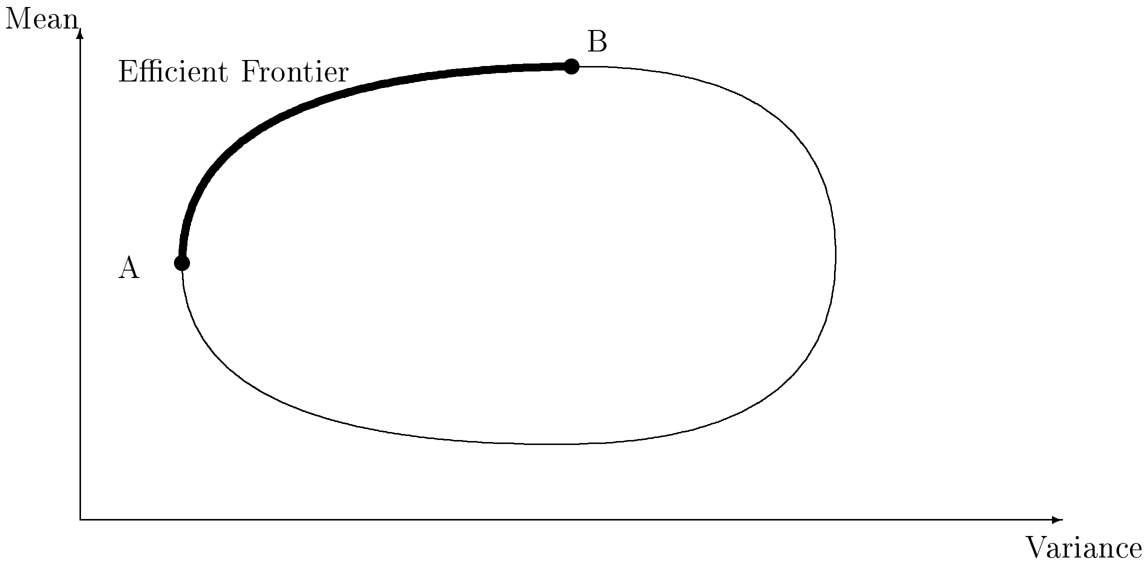

Figure 2.3: The efficient frontier

The utility function and its derivatives are therefore given by

$$U(\tilde{x}) = \tilde{x} + b\tilde{x}^2, \tag{2.8}$$

$$U'(\tilde{x}) = 1 + 2b\tilde{x}, \tag{2.9}$$

$$U''(\tilde{x}) = 2b. \tag{2.10}$$

According to (2.5) the Arrow-Pratt measure of risk aversion turns out to be

$$z = -\frac{2b}{1 + 2bE[\tilde{x}]}. \tag{2.11}$$

If we concentrate on risk averse individuals and assume reasonably positive marginal utility, (2.11) implies that

$$b < 0. \tag{2.12}$$

But if $b < 0$ we see from (2.9) that the marginal utility is only positive if

$$E[\tilde{x}] < -\frac{1}{2b}. \tag{2.13}$$

For large expected values the marginal utility can become negative. This unreasonable result can only be ruled out if the risk aversion is sufficiently small.[28]

[28] This restriction on the expected value is another argument often used against the use of

If we define $E[\tilde{x}] = \mu$ and $Var[\tilde{x}] = \sigma^2$ we can write expected utility as

$$E[U(\tilde{x})] = E\left[\tilde{x} + b\tilde{x}^2\right] = \mu + bE\left[\tilde{x}^2\right] = \mu + b\left(\mu^2 + \sigma^2\right). \tag{2.14}$$

The indifference curves are obtained by totally differentiating both sides:

$$dE[U(\tilde{x})] = (1 + 2b\mu)d\mu + 2\sigma d\sigma = 0. \tag{2.15}$$

The slope of the indifference curve in the (μ,σ)-plane is obtained by rearranging (2.15):[29]

$$\frac{d\mu}{d\sigma} = -\frac{2b\sigma}{1 + 2b\mu} = z\sigma > 0, \tag{2.16}$$

i.e. for risk averse investors the indifference curves have a positive slope in the (μ,σ)-plane.

The equation of the indifference curve is obtained by solving (2.14) for μ:[30]

$$\begin{aligned} E[U(\tilde{x})] &= \mu + b\mu^2 + b\sigma^2 \\ \mu^2 + \frac{1}{b}\mu + \sigma^2 &= \frac{E[U(\tilde{x})]}{b} \\ \left(\mu + \frac{1}{2b}\right)^2 + \sigma^2 &= \frac{1}{b}E[U(\tilde{x})] + \frac{1}{4b^2}. \end{aligned} \tag{2.17}$$

Defining $r^* = -\frac{1}{2b}$ as the expected outcome that must not be exceeded for the marginal utility to be positive according to equation (2.13), we can rewrite the equation for the indifference curves as

$$(\mu - r^*) + \sigma^2 = -2r^*E[U(\tilde{x})] + r^{*2}, \tag{2.18}$$

a quadratic utility function and hence the mean-variance criterion. Another argument is that the risk aversion increases with the expected outcome: $\frac{\partial z}{\partial E[\tilde{x}]} = \frac{4b^2}{(1+2bE[\tilde{x}])^2} > 0$, which contradicts empirical findings. Moreover in many theoretical models a constant risk aversion is assumed, which has been shown by PRATT (1964) to imply an exponential utility function. If the expected outcome does not vary too much constant risk aversion can be approximated by using a quadratic utility function.

[29] Instead of using the variance as a measure of risk, it is more common to use its square root, the standard deviation. As the square root is a monotone transformation the results are not changed by this manipulation.

[30] See SHARPE (1970, pp. 198 f.)

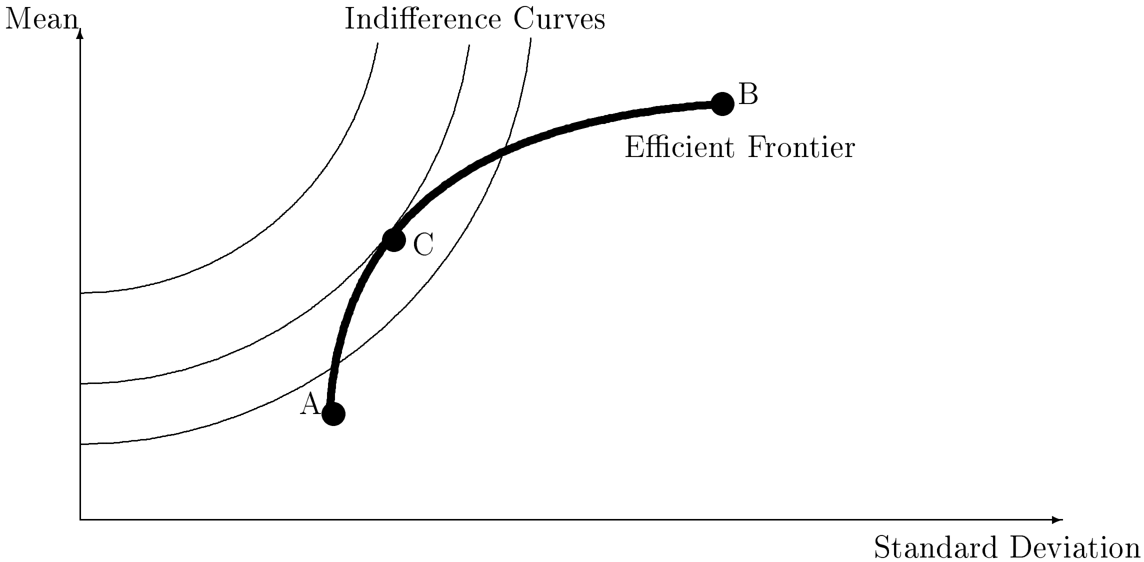

Figure 2.4: Determination of the optimal alternative

which is the equation of a circle with center $\mu = r^*$, $\sigma = 0$ and radius $\sqrt{-2r^* E\left[U\left(\tilde{x}\right)\right] + r^{*2}}$.[31] With this indifference curve, which has as the only parameter a term linked to the risk aversion, it is now possible to determine the optimal alternative out of the efficient alternatives, that is located at the point where the efficient frontier is tangential to the indifference curve. Figure (2.4) shows the determination of the optimal alternative C.

We will now show that with a quadratic utility function the mean-variance criterion is optimal.[32] We assume two alternatives with $\mu_i = E\left[a_i\right] > E\left[a_j\right] = \mu_j$. Let further $\sigma_i^2 = Var\left[a_i\right]$ and $\sigma_j^2 = Var\left[a_j\right]$. If $a_i \succ a_j$ it has to be shown that

$$E\left[U\left(a_i\right)\right] > E\left[U\left(a_j\right)\right].$$

Substituting the utility functions gives

$$\mu_i + b\mu_i^2 + b\sigma_i^2 > \mu_j + b\mu_j^2 + b\sigma_j^2$$

[31] The results that the indifference curves are circles gives rise to another objection against the use of a quadratic utility function. An individual with a quadratic utility function should be indifferent between an expected outcome of $r^* + v$ and $r^* - v$ for any given σ. From the mean-variance criterion (2.7) we know that for a given σ the alternative with the higher expected outcome will be preffered. In practice this problem is overcome by using only the lower right sector of the circle.

[32] Such a proof is given e.g. in Levy/Sarnat (1972, pp. 385 ff.).

$$\mu_i - \mu_j + b(\mu_i^2 - \mu_j^2) + b(\sigma_i^2 - \sigma_j^2) = (\mu_i - \mu_j)\left[1 + b(\mu_i + \mu_j)\right] + b(\sigma_i^2 - \sigma_j^2) > 0.$$

Dividing by $-2b > 0$ gives us

$$(\mu_i - \mu_j)\left[-\frac{1}{2b} - \frac{\mu_i + \mu_j}{2}\right] - \frac{\sigma_i^2 - \sigma_j^2}{2} > 0. \tag{2.19}$$

From (2.13) we know that $-\frac{1}{2b} > \mu_i$ and $-\frac{1}{2b} > \mu_j$, hence we find that

$$-\frac{1}{2b} > \frac{\mu_i + \mu_j}{2} \tag{2.20}$$

With the assumption that $\mu_i > \mu_j$ and as $b < 0$ the first term in (2.19) is positive. If now $\sigma_i^2 \leq \sigma_j^2$ as proposed by the mean-variance criterion equation (2.19) is fulfilled and we have shown that the it represents the true preferences.

If $\sigma_i^2 > \sigma_j^2$ in general nothing can be said which alternative will be preferred. For $\mu_i = \mu_j$ we need $\sigma_i^2 < \sigma_j^2$ in order to prefer a_i over a_j. This is exact the statement made by the mean-variance criterion in (2.13). Therewith it has been shown that in the case of a quadratic utility function the mean-variance criterion is optimal, i.e. represents the true preferences.[33]

2.4 The Markowitz Frontier

The portfolio selection theory is based on several assumptions that are summarized in table 2.1 in the form they have been stated by Lintner (1965a, p. 15). Some of these assumptions, like the absence of transaction costs and taxes have been lifted by more recent contributions without giving fundamental new insights.

In portfolio selection theory the different alternatives are the compositions of the portfolios, i.e. the weight each asset has.[34] Assume an investor has to choose between $N > 1$ assets, assigning a weight of x_i $(i = 1, \ldots, N)$ to each asset.

[33] A quadratic utility function is not only a sufficient condition for the optimality of the mean-variance criterion, but also a necessary condition. This is known in literature as the Schneeweiss-Theorem.

[34] The decision which portfolio is optimal does not depend on total wealth for a given constant risk aversion, hence it can be analyzed by dealing with shares only. See Levy/Sarnat (1972, pp. 420 f.).

• No transaction costs and taxes
• Assets are indefinitely divisible
• Each investor can invest into every asset without restrictions
• Investors maximize expected utility by using the mean-variance criterion
• Prices are given and cannot be influenced by investors (competitive prices)
• The model is static, i.e. only a single time period is considered

Table 2.1: Assumptions of the portfolio theory*Source: After Lintner (1965), p. 15*

The expected return (mean) of each asset is denoted μ_i $(1 = 1, \ldots, N)$ and the variance of the returns by $\sigma_i^2 > 0$ $(i = 1, \ldots, N)$.[35] The covariances between assets i and j will be denoted σ_{ij} $(i, j = 1, \ldots, N)$.

The weight of an asset an investor holds have to sum up to one and are assumed to be positive as we do not allow for short sales at this stage:

$$\sum_{i=1}^{N} x_i = 1 \qquad (2.21)$$
$$x_i \geq 0, \; i = 1, \ldots, N$$

For the moment assume that there are only $N = 2$ assets. The characteristics of each asset can be represented as a point in the (μ,σ)-plane. We then can derive the location of any portfolio in the (μ,σ)-plane by combining these two assets.[36]

The expected return and the variance of the return of the portfolio is given by

$$\mu_p = x_1\mu_1 + x_2\mu_2 = \mu_2 + x_1(\mu_1 - \mu_2), \qquad (2.22)$$

$$\begin{aligned} \sigma_p^2 &= x_1^2\sigma_1^2 + x_2^2\sigma_2^2 + 2x_1x_2\sigma_{12} \\ &= \sigma_2^2 + x_1^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}) + 2x_1(\sigma_1\sigma_2\rho_{12} - \sigma_2^2), \end{aligned} \qquad (2.23)$$

[35] Instead of investigating final expected wealth and its variance after a given period of time (the time horizon) we can use the expected return and variances of returns as they are a positive linear transformation of the wealth. As has been noted above the decision is not influenced by such transformation.

[36] See Tobin (1966, pp. 22ff.).

where $\rho_{12} = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$ denotes the correlation of the two assets.

The portfolio with the lowest risk is obtained by minimizing (2.23). The first order condition is

$$\frac{\partial \sigma_p^2}{\partial x_1} = 2x_1(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}) + 2(\sigma_1\sigma_2\rho_{12} - \sigma_2^2) = 0.$$

The second order condition for a minimum is fulfilled if $\sigma_1 \neq \sigma_2$ or if $\sigma_1 = \sigma_2$ and $\rho_{12} \neq 1$:

$$\frac{\partial^2 \sigma_p^2}{\partial x_1^2} = 2(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}) > 2(\sigma_1 - \sigma_2)^2 > 0$$

Solving the first order condition gives the weights for the *minimum risk portfolio* (*MRP*):

$$x_1^{MRP} = \frac{\sigma_2^2 - \sigma_1\sigma_2\rho_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}. \tag{2.24}$$

The minimum variance can be obtained by inserting (2.24) into (2.23):

$$\begin{aligned} \sigma_{MRP}^2 &= \sigma_2^2 + \frac{(\sigma_2^2 - \sigma_1\sigma_2\rho_{12})^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}} - 2\frac{(\sigma_2^2 - \sigma_1\sigma_2\rho_{12})^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}} \\ &= \sigma_2^2 - \frac{(\sigma_2^2 - \sigma_1\sigma_2\rho_{12})^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}} \\ &= \frac{\sigma_1^2\sigma_2^2(1 - \rho_{12}^2)}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}. \end{aligned} \tag{2.25}$$

If the returns of the two assets are uncorrelated ($\rho_{12} = 0$), then (2.25) reduces to

$$\sigma_{MRP}^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \tag{2.26}$$

This variance is smaller than the variance of any of these two assets.[37] By holding an appropriate portfolio, the variance, and hence the risk, can be reduced, whereas the expected return lies between the expected returns of the two assets.

With perfectly negative correlated assets ($\rho_{12} = -1$) we find that

$$\sigma_{MRP}^2 = 0 \tag{2.27}$$

[37] Suppose $\sigma_{MRP}^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} > \sigma_2^2$, this would imply that $\sigma_1^2 > \sigma_1^2 + \sigma_2^2$ and hence $\sigma_2^2 < 0$, which contradicts the assumption that $\sigma_2^2 > 0$. A similar argument can be used to show that $\sigma_{MRP}^2 < \sigma_1^2$.

and the risk can be eliminated from the portfolio.

In the case of perfectly correlated assets ($\rho_{12} = 1$) the minimum variance is the variance of the asset with the lower variance:

$$\sigma^2_{MRP} = \begin{cases} \sigma_1^2 & \text{if} \quad \sigma_1^2 \leq \sigma_2^2 \\ \sigma_2^2 & \text{if} \quad \sigma_1^2 > \sigma_2^2 \end{cases} . \tag{2.28}$$

We can derive a general expression for the mean-variance relation:

$$\begin{aligned} \sigma_p^2 - \sigma^2_{MRP} &= \sigma_2^2 + x_1^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}) \\ &\quad +2x_1(\sigma_1\sigma_2\rho_{12} - \sigma_2^2) \\ &\quad -\frac{\sigma_1^2\sigma_2^2(1-\rho_{12}^2)}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}} \\ &= x_1^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}) \\ &\quad -2x_1x_1^{MRP}(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}) \\ &\quad +\frac{\sigma_1^2\sigma_2^2 + \sigma_2^4 - 2\sigma_1\sigma_2^3\rho_{12} - \sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_2^2\rho_{12}^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}} \\ &= x_1^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}) \\ &\quad -2x_1x_1^{MRP}(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}) \\ &\quad +\frac{(\sigma_2^2 - \sigma_1\sigma_2\rho_{12})^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}} \\ &= (x_1^2 - 2x_1x_1^{MRP} + (x_1^{MRP})^2)(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}) \\ &= (x_1 - x_1^{MRP})^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}). \end{aligned} \tag{2.29}$$

With μ_{MRP} denoting the expected return of the minimum risk portfolio, we find that

$$\mu_p - \mu_{MRP} = (x_1 - x_1^{MRP})(\mu_1 - \mu_2). \tag{2.30}$$

Solving (2.30) for $x_1 - x_1^{MRP}$ and inserting into (2.29) we obtain:

$$\sigma_p^2 - \sigma^2_{MRP} = (\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12})\frac{(\mu_p - \mu_{MRP})^2}{(\mu_1 - \mu_2)^2}.$$

Rearranging gives

$$(\mu_p - \mu_{MPR})^2 = \frac{\sigma_p^2 - \sigma_{MRP}^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}(\mu_1 - \mu_2)^2. \tag{2.31}$$

This equation represents a hyperbola with axes[38]

$$\begin{aligned} \mu_p &= \mu_{MRP} + \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}}\sigma_p, \\ \mu_p &= \mu_{MRP} - \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}}\sigma_p. \end{aligned}$$

The efficient portfolios lie on the upper branch of this hyperbola, i.e. above the minimum risk portfolio.[39] Figure 2.5 shows the efficient portfolios for different correlations. It can easily be shown that in the case of perfect positive correlation the efficient portfolios are located on a straight line connecting the two assets, in case of perfectly negative correlation on straight lines connecting the assets with the minimum risk portfolio. Between efficient portfolios can only be distinguished by using the utility function. Figure 2.6 adds the indifference curve to the opportunity locus and determines the location of the *optimal portfolio* (OP). The location of the optimal portfolio depends on the risk aversion of the investor, the more risk averse the investor is the more close the optimal portfolio will be located to the minimum risk portfolio.

It now possible to introduce a third asset. In a similar way hyperbolas can be deducted representing all combinations of this asset with one of the other two. Furthermore we can view any portfolio consisting of the two other assets as a single new asset and can combine it in the same manner with the third asset. Figure 2.7 illustrates this situation. All achievable portfolio combinations are now located in the gray shaded area, where the bold line encircling the different hyperbolas is the new opportunity locus.

This concept can be generalized to $N > 3$ assets in the same manner. All achievable assets will be located in an area and the efficient frontier will be a

[38] See TOBIN (1966, p.30).
[39] The efficient frontier is also called *opportunity locus*.

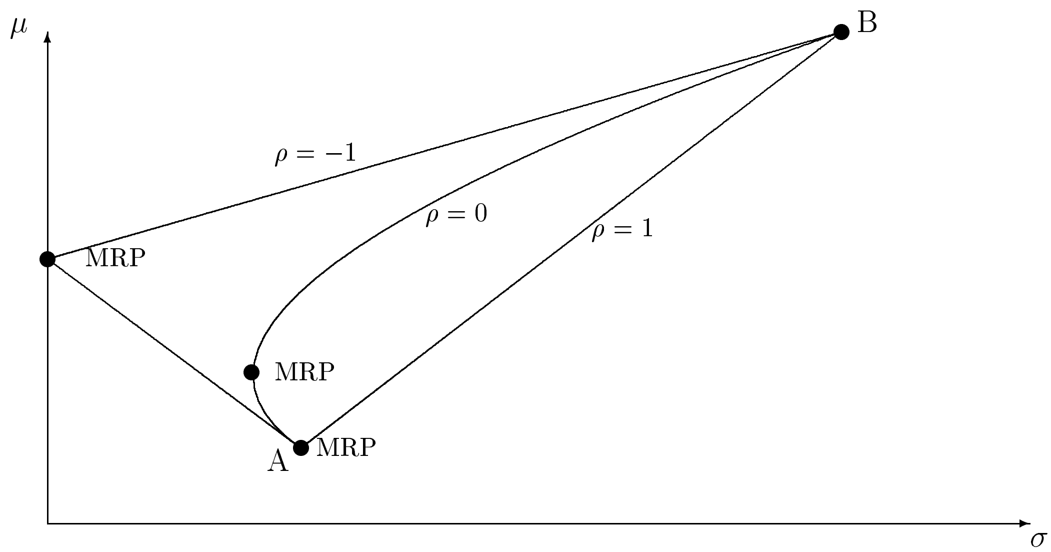

Figure 2.5: Efficient Portfolios with two assets

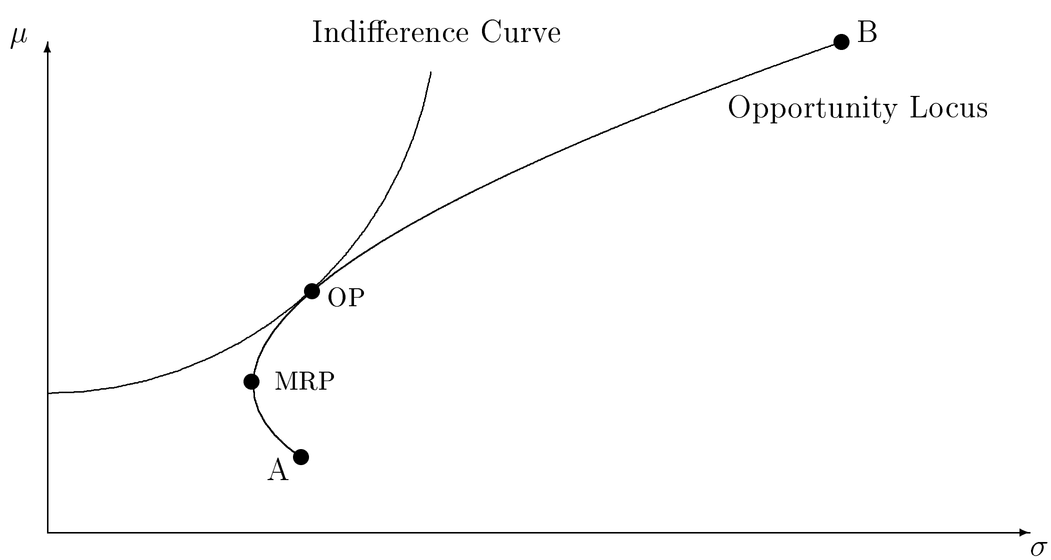

Figure 2.6: Determination of the optimal portfolio with two assets

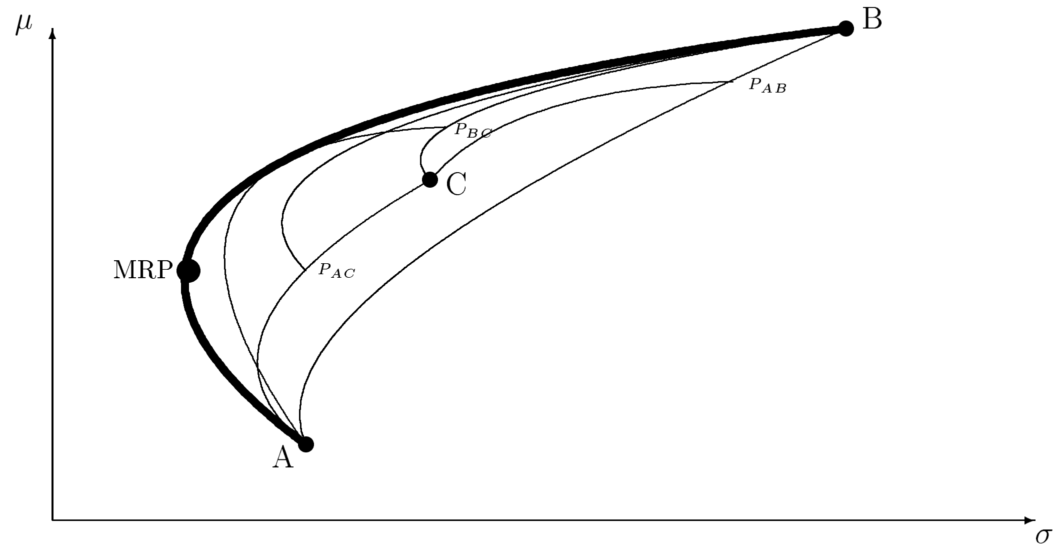

Figure 2.7: Portfolio Selection with three assets

hyperbola. Using the utility function the optimal portfolio can be determined in a similar way as in the case of two assets as shown in figure 2.8. If an asset is added the area of achievable portfolios is enlarged and encompasses the initial area. This can simply be proofed by stating that the new achievable portfolios encompass also the portfolios assigning a weight of zero to the new asset. With a weight of zero these portfolios are identical to the initially achievable portfolios. To these portfolios those have to be added assigning a non-zero weight to the new asset. Therefore the efficient frontier moves further outward to the upper left. By adding new assets the utility can be increased.

Thus far it has been assumed that $\sigma_i^2 > 0$, i.e. all assets were risky. It is also possible to introduce a riskless asset, e.g. a government bond, with a variance of zero. Define the return of the riskless asset by r, then in the two asset case we get from (2.22) and (2.23):[40]

$$\mu_p = x_1\mu_1 + x_2 r = r + x_1(\mu_1 - r), \tag{2.32}$$

$$\sigma_p^2 = x_1^2\sigma_1^2. \tag{2.33}$$

[40] See TOBIN (1958).

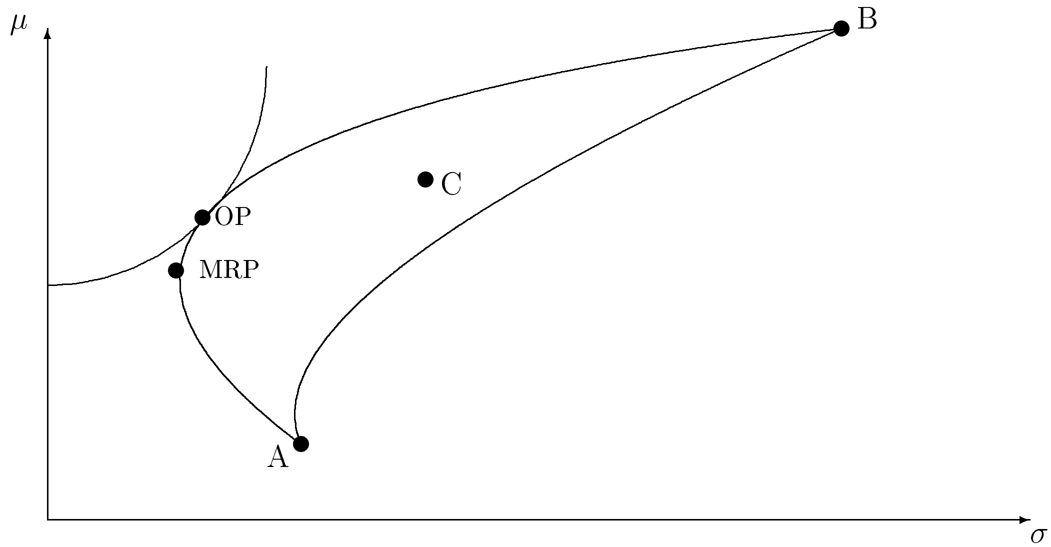

Figure 2.8: The optimal portfolio with $N > 2$ assets

Solving (2.33) for x_1 and inserting into (2.32) gives

$$\mu_p = r + \frac{\sigma_p}{\sigma_1}(\mu_1 - r) = r + \frac{\mu_1 - r}{\sigma_1}\sigma_p. \tag{2.34}$$

The expected return of the portfolio is linear in the variance of the portfolio return, i.e. the hyperbola reduces to a straight line from the location of the riskless asset, (0, r), to the location of the asset. In the case of many risky assets we can combine every portfolio of risky assets with the riskless asset and obtain all achievable portfolios. As shown in figure 2.9 all achievable portfolios are located between the two straight lines, the upper representing the efficient frontier. There exists a portfolio consisting only of risky assets that is located on the efficient frontier. It is the portfolio consisting only of the risky assets at which the efficient frontiers with and without a riskless asset are tangential.[41] This portfolio is called the *optimal risky portfolio* (*ORP*). The efficient frontier with a riskless asset is also called the *capital market line.*

All efficient portfolios are located on the capital market line, consequently they are a combination of the riskless asset and the optimal risky portfolio. The

[41] It is also possible that no tangential point exists, in this case a boundary solution exists and the risky portfolio consists only of a single risky asset.

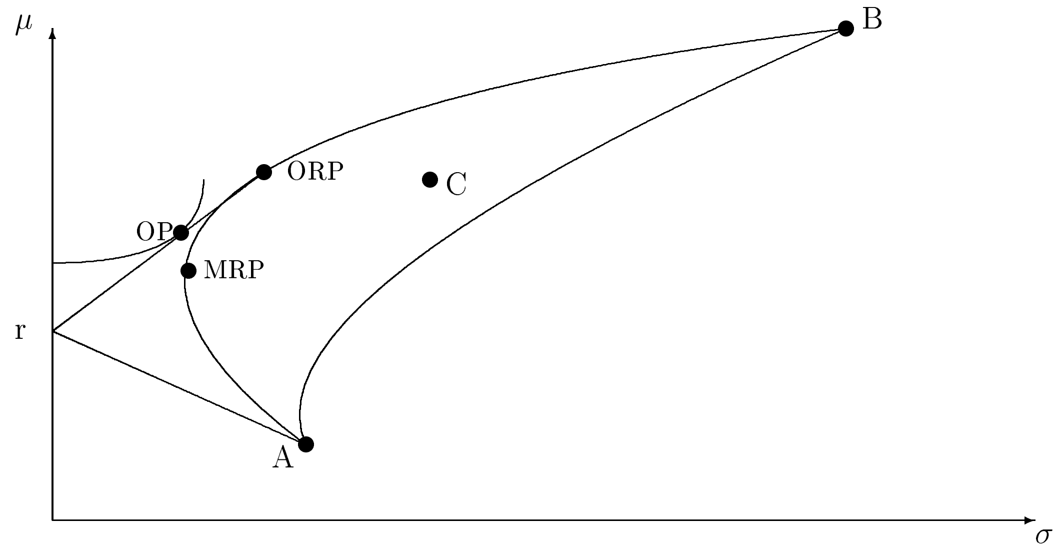

Figure 2.9: The optimal Portfolio with a riskless asset

optimal portfolio can be obtained in the usual way by introducing the indifference curves. As the optimal portfolio always is located on the capital market line, it consists of the risky asset and the optimal risky portfolio. Which weight is assigned to each depends on the risk aversion of the investor, the more risk averse he is the more weight he will put on the riskless asset. The weights of the optimal risky portfolio do not depend on the risk aversion of the investor. The decision process can therefore be separated into two steps, the determination of the optimal risky portfolio and then the determination of the optimal portfolio as a combination of the ORP with the riskless asset. As this result has first been presented by TOBIN (1958) it is also called the *Tobin separation theorem.*[42]

So far we have assumed that for all $i = 1, \ldots, N : x_i \geq 0$. If we allow now some x_i to be negative the possibilities to form portfolios is extended. An asset with $x_i < 0$ means that the asset is sold short, i.e. it is sold without having owned it before. This situation can be viewed as a credit that has not been given and has

[42] For investors being less risk averse it is possible that the optimal portfolio is located on the part of the efficient frontier above the ORP, in this case the optimal portfolio does not contain the riskless asset and assigns different wights to the risky assets compared to the ORP. Therefore in general the Tobin separation theorem does only hold with the inclusion of short sales, as described in the next paragraph.

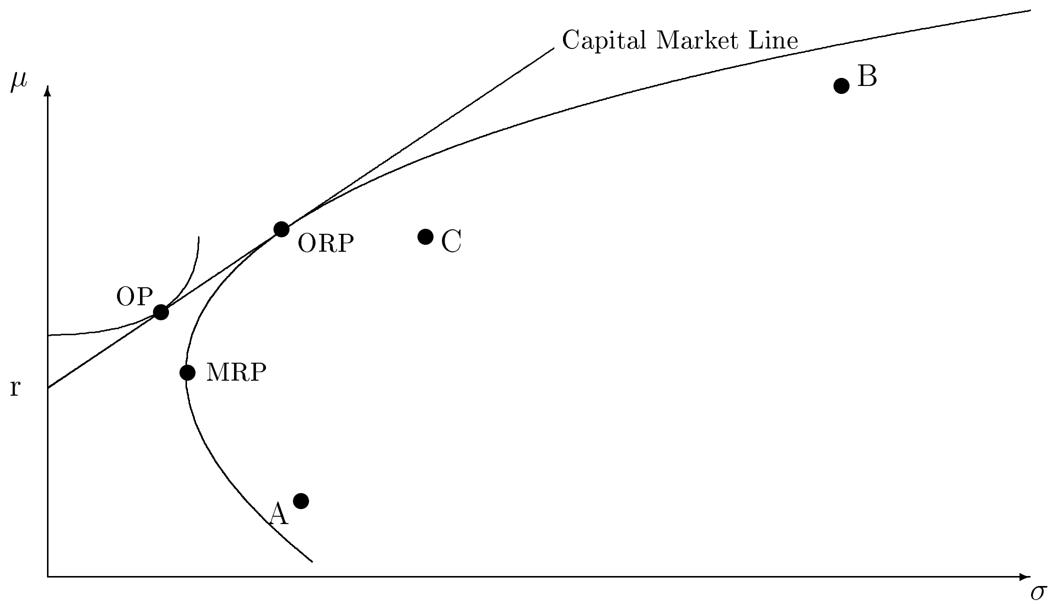

Figure 2.10: Portfolio Selection with Short Sales

to be repaid in money (unless money is the asset), but in the asset. The assets can be the riskless asset or the risky assets, in the former case the short sale is an ordinary credit. It is assumed that credits can be obtained at the same condition (interest rate or expected return and risk) as investing in the asset.

By allowing short sales the efficient frontier of the risky portfolios further moves to the upper left as new possibilities to form portfolios are added by lifting the restriction that the weights must be non-negative. Therewith the capital market line becomes steeper and the utility of the optimal portfolio increases. Figure 2.10 illustrates this case. The Capital Market Line extends beyond the ORP and therewith the optimal portfolio will always be a combination of the riskless asset and the ORP. The Tobin separation theorem applies in all cases, independent of the degree of risk aversion. If the ORP is located above the ORP the riskless asset is sold short and a higher fraction of the optimal portfolio consists of the ORP.

In applying the portfolio theory to determine the optimal portfolio several problems are faced:

- Determination of the risk aversion of the investor
- Determination of the expected returns, variances and covariances of the assets
- Computation of the efficient frontier and the optimal portfolio

There exists no objective way to determine the risk aversion of an investor, most investors are only able to give a qualitative measure of their risk aversion, if at all. The transformation into a quantitative measure is an unsolved, but for the determination of the optimal portfolio critical problem. It is important for the allocation between the riskless asset and the optimal risky portfolio.

Expected returns, variances and covariances can be obtained from estimates based on past data. But there is no guarantee that these results are reasonable for the future. It is also possible to use other methods to determine these moments, e.g. by using subjective beliefs. The determination of these moments are critical for the determination of the optimal risky portfolio.

The computational problem to determine the efficient frontier and the optimal portfolio requires to apply non-trivial numerical optimization routines.[43] Advances in computer facilities and the availability of these routines do not impose a threat anymore as it has done in former years.

When having solved the above mentioned problems, the portfolio theory does allow to answer the questions raised at the beginning of this section:

- the share to be invested into risky assets, i.e. in our case the stock market is determined by the optimal portfolio,
- the assets to invest in are those included in the optimal risky portfolio,

[43] For a detailed description of the mathematical concepts to solve these problems see Markowitz (1959) and Aschinger (1990).

- the shares to invest in each selected asset are given by the weights of the optimal risky portfolio.

The portfolio theory has developed a theory how to allocate resources optimal. Although mostly only financial assets are included, other assets like human capital, real estate and others can easily be included, although it is even more difficult to determine their characteristics.

A shortcoming of the portfolio theory is that it is a static model. It determines the optimal portfolio at a given date. If the time horizon is longer than one period, the prices of assets change over time, and therewith the weights of the assets in the initial portfolio change. Even if the expected returns, variances and covariances do not change this requires to rebalance the portfolio every period. As assets with a high realized return enlarge their weight, they have partially to be sold to buy assets which had a low return (sell the winners, buy the losers). In a dynamic model other strategies have been shown to achieve a higher expected utility for investors, but due to the static nature of the model such strategies cannot be included within this framework.

3. The Capital Asset Pricing Model

In this section we will derive the *Capital Asset Pricing Model* (*CAPM*) that is the most prominent model in asset pricing. Sharpe (1964) and Lintner (1965a) developed the CAPM independent of each other by using the portfolio theory to establish a market equilibrium.

3.1 Derivation of the CAPM

The basis of the CAPM is the portfolio theory with a riskless asset and unlimited short sales. We do not consider only the decision of a single investor, but aggregate them to determine a market equilibrium. In portfolio theory the price of an asset was exogeneously given and could not be influenced by any investor. Given this price he formed his beliefs on the probability distribution.[1] The beliefs were allowed to vary between investors.

In this section asset prices (or equivalently expected asset returns) will no longer be exogenously given, but be an equilibrium of the market. Recalling the results from section 4.1 we know that the current price affects the expected returns and vice versa. Given future expected dividends and assuming that markets are efficient, i.e. that the prices of assets equal their fundamental value, a high current price results in a low expected return in the next period and a low current price in a high expected return. In the same way in order to expect a high return, the price has to be low and for a low expected return a high price is needed. This equivalence of price and return allows us to concentrate either on prices or on

[1] As portfolio theory makes use of the mean-variance criterion it is sufficient to form beliefs only about means, variances and covariances instead of the entire distribution.

expected returns. Convention in the academic literature requires us to focus on expected returns.

Additionally to the assumptions already stated for portfolio theory, we have to add that all investors have the same beliefs on the probability distribution of all assets, i.e. agree on the expected returns, variances and covariances.[2] If all investors agree on the characteristics of an asset the optimal risky portfolio will be equal for all investors, even if they differ in their preferences (risk aversion). Because all assets have to be held by the investors the share each asset has in the optimal risky portfolio has to be equal to its share of the market value of all assets.[3] The optimal risky portfolio has to be the market portfolio. Moreover all assets have to be marketable, i.e. all assets must be traded and there are no other investment opportunities not included into the model. Table 3.1 summarizes the assumptions.

Every investor j $(j = 1, \ldots, M)$ maximizes his expected utility by choosing an optimal portfolio, i.e. choosing optimal weights for each asset. With the results of section 4.1.2 we get[4]

$$\begin{aligned} \max_{\{x_i\}_{i=1}^N} E\left[U^j(R_p)\right] &= \max_{\{x_i\}_{i=1}^N} U^j\left(\mu_p - \frac{1}{2} z_j \sigma_p^2\right) \\ &= \max_{\{x_i\}_{i=1}^N} \left(\mu_p - \frac{1}{2} z_j \sigma_p^2\right) \\ &= \max_{\{x_i\}_{i=1}^N} \left(\sum_{i=1}^N x_i \mu_i - \frac{1}{2} z_j \sum_{k=1}^N \sum_{i=1}^N x_i x_k \sigma_{ik}\right) \end{aligned} \tag{3.1}$$

for all $j = 1, \ldots, M$ with the restriction $\sum_{i=1}^N x_i = 1$.

[2] See Sharpe (1964, p. 433), Lintner (1965a, p. 600) and Lintner (1965b, p.25). Lintner (1965b) calls this assumption *idealized uncertainty*. Sharpe (1970, pp. 104 ff.) also considers different beliefs. As the main line of argument does not change, these complications are not further considered here. Several assumptions made in portfolio theory can also be lifted without changing the results significantly. Black (1972) restricts short sales and Sharpe (1970, pp. 110 ff.) applies different interest rates for borrowing and lending the riskless asset.

[3] See Sharpe (1970, p. 82).

[4] See Dumas/Allaz (1996, pp. 61 f. and pp. 78 f.).

• No transaction costs and taxes
• Assets are indefinitely dividable
• Each investor can invest into every asset without restrictions
• Investors maximize expected utility by using the mean-variance criterion
• Prices are given and cannot be influenced by the investors (competitive prices)
• The model is static, i.e. only a single time period is considered
• Unlimited short sales
• Homogeneity of beliefs
• All assets are marketable

Table 3.1: Assumptions of the CAPM

The Lagrange function for solving this problem can easily be obtained as

$$L_j = \sum_{i=1}^{N} x_i \mu_i - \frac{1}{2} z_j \sum_{k=1}^{N} \sum_{i=1}^{N} x_i x_k \sigma_{ik} + \lambda \left(1 - \sum_{i=1}^{N} x_i \right). \tag{3.2}$$

The first order conditions for a maximum are given by

$$\frac{\partial L_j}{\partial x_i} = \mu_i - z_j \sum_{k=1}^{N} x_k \sigma_{ik} - \lambda = 0, \qquad i = 1, \ldots, N, \tag{3.3}$$

$$\frac{\partial L_j}{\partial \lambda} = 1 - \sum_{i=1}^{N} x_i = 0 \tag{3.4}$$

for all $j = 1, \ldots, M$.[5]

Solving the above equations for μ_i gives

$$\begin{aligned} \mu_i &= \lambda + a_j \sum_{k=1}^{N} x_k \sigma_{ik} = \lambda + z_j Cov \left[R_i, \sum_{k=1}^{N} x_k R_k \right] \\ &= \lambda + z_j Cov \left[R_i, R_p \right] = \lambda + z_j \sigma_{ip}. \end{aligned} \tag{3.5}$$

[5] The second order condition for a maximum can be shown to be fulfilled, due to space limitations this proof is not presented here.

With $\sigma_{ip} = 0$ we find that $\mu_i = \lambda$, hence we can interpret λ as the expected return of an asset which is uncorrelated with the market portfolio. As the riskless asset is uncorrelated with any portfolio, we can interpret λ as the risk free rate of return r:

$$\mu_i = r + z_j \sigma_{ip}. \tag{3.6}$$

From (3.6) we see that the expected return depends linearly on the covariance of the asset with the market portfolio. The covariance can be interpreted as a measure of risk for an individual asset (*covariance risk*). Initially we used the variance as a measure of risk, but as has been shown in the last section the risk of an individual asset can be reduced by holding a portfolio. The risk that cannot be reduced further by diversification is called *systematic risk*, whereas the diversifiable risk is called *unsystematic risk*. The total risk of an asset consists of the variation of the market as a whole (systematic risk) and an asset specific risk (unsystematic risk). As the unsystematic risk can be avoided by diversification it is not compensated by the market, efficient portfolios therefore only have systematic and no unsystematic risk.[6]

The covariance of an asset can be also interpreted as the part of the systematic risk that arises from an individual asset:

$$\begin{aligned} \sum_{i=1}^{N} x_i \sigma_{ip} &= \sum_{i=1}^{N} x_i Cov[R_i, R_p] = Cov\left[\sum_{i=1}^{N} x_i R_i, R_p\right] \\ &= Cov[R_p, R_p] = Var[R_p] = \sigma_p^2. \end{aligned}$$

Equation (3.6) is valid for all assets and hence for any portfolio, as the equation for a portfolio can be obtained by multiplying with the appropriate weights and then summing them up, so that we can apply this equation also to the market portfolio, which is also the optimal risky portfolio:

$$\mu_p = r + z_j \sigma_p^2. \tag{3.7}$$

[6] See SHARPE (1964, pp. 438 f.) and SHARPE (1970, p.97).

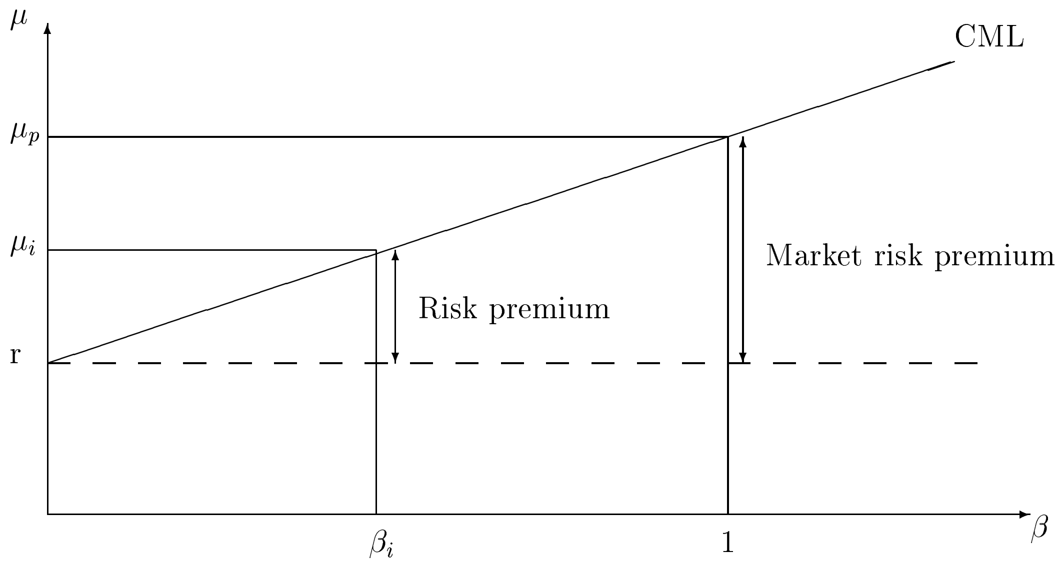

Figure 3.1: The Capital Asset Pricing Model

Solving for z_j and inserting into (3.6) gives us the usual formulation of the CAPM:

$$\mu_i = r + (\mu_p - r)\frac{\sigma_{ip}}{\sigma_p^2}. \tag{3.8}$$

Defining $\beta_i = \frac{\sigma_{ip}}{\sigma_p^2}$ we can rewrite (3.8) as

$$\mu_i = r + (\mu_p - r)\beta_i. \tag{3.9}$$ [7]

β_i represents the relative risk of the asset i (σ_{ip}) to the market risk (σ_p^2). The beta for the market portfolio is easily shown to be 1. We find a linear relation between the expected return and the relative risk of an asset. This relation is independent of the preferences of the investors (z_j), provided that the mean-variance criterion is applied and that the utility function is quadratic. This equilibrium line is called the *Capital Market Line* (*CML*). Figure 3.1 illustrates this relation.

For the risk an investor takes he is compensated by the amount of $\mu_p - r$ per unit of risk, the total amount $(\mu_p - r)\beta_i$ is called the *risk premium* or the *market price of risk*. The risk free rate of return r may be interpreted as the price for time.

[7] It also can frequently be found that only excess returns over the risk free rate of return are considered. by defining $\mu'_j = \mu_j - r$ (3.9) becomes $\mu'_i = \mu'_p\beta_i$.

It is the compensation for not consuming the amount in the current period, but wait until the next period.[8]

Equation (3.9) presents a formula for the expected return given the interest rate, r, the beta and the expected return on the market portfolio, μ_p. However, the expected return of the market portfolio is not exogenous as it is a weighted average of the expected returns of the individual assets. In this formulation only relative expected returns can be determined, the level of expected returns, i.e. the market risk premium, is not determined by the CAPM. Although we can reasonably assume μ_p to be given when investigating a small capitalized asset, we have to determine μ_p endogenously.

Thus far we only considered the market portfolio, but an investor will in general not hold the market portfolio (optimal risky portfolio). As we saw in section 4.2 the optimal portfolio will be a combination of the market portfolio and the riskless asset, where the shares will vary among investors. We can now add another restriction to our model. The return on the optimal risky portfolio has to be such that the market for the riskless assets has to be in equilibrium. The amounts of riskless assets lent and borrowed have to be equal:[9]

$$\sum_{j=1}^{M} x_{jr} = 0, \tag{3.10}$$

where x_{jr} denotes the demand of the jth investor for the riskless asset. With this additional market to be in equilibrium it is possible to determine μ_p endogenously. We herewith have found an equilibrium in the expected returns. These expected returns can now be used to determine equilibrium prices according to section 4.1.

Nevertheless this result remains to determine only relative prices. The risk free rate r is not determined endogenously. Although is can reasonably be assumed

[8] See SHARPE (1964, p. 425). This in opposition to the interpretation of KEYNES (1997, pp. 165) who interpreted the interest rate as a compensation for giving up liquidity, i.e. binding resources over time.

[9] See SHARPE (1970, p.81). Also a fixed supply of the riskless asset can be assumed without changing the argument.

that r can substantially be influenced by monetary policy, especially for longer time horizons it is not given.

3.2 Critique of the CAPM

The assumptions underlying the CAPM are very restrictive. The problems associated with the use of the mean-variance criterion and the quadratic utility function have already been mentioned in section 4.2. Some restrictions such as the absence of transaction costs and taxes, unlimited borrowing and lending at the risk free rate and short sales have been lifted by more recent contributions without changing the results significantly. One restriction however that mostly is not mentioned in the literature is that the assets have to be linearly dependent. This linearity, also used in portfolio theory is implied by the use of the covariance, which is only able to capture linear dependencies appropriately. This linearity of returns rules out the inclusion of derivatives that mostly have strong non-linearities in their pay-offs and have become an important tool for investment in recent years. By excluding such assets the need to include all assets is violated.

Beside these theoretical critiques empirical investigations show a mixed support for the CAPM. There exist a large number of empirical investigations of the CAPM using different econometric specifications.[10] Early investigations mainly supported the CAPM,[11] but more recent results show that the CAPM is not able to explain the observed returns. If other variables, such as book-market ratios, market value of a company or price-earnings ratios are included the beta has no significant influence on the observed returns.[12]

Like the portfolio theory the CAPM is a static model, i.e. the investment horizon

[10] See Campbell et al. (1997, pp. 211 ff.) and Dumas/Allaz (1996, pp. 128) for an overview of the different econometric models.
[11] See e.g. Fama/MacBeth (1973).
[12] See Fama/French (1992) and Fama/French (1993).

is assumed to be only a single period. As has been pointed out, after every period the portfolio has to be rebalanced even if the beliefs do not change. This rebalancing will affect the market equilibrium prices and hence the expected future returns and risk premia. Therefore a dynamic model will capture the nature of decision making more appropriate, such a modification is given with the Intertemporal CAPM to be presented in section 4.7.

Another critical point in the CAPM is that unconditional beliefs (means, variances and covariances) are used, the investors are not able to condition their beliefs on information they receive. A direct implication of this is the assumption that beliefs are constant over time. Many empirical investigations give strong support that beliefs are varying over time, of special importance is the beta. But also the risk aversion and the risk premium have been found to vary significantly over time. These aspects are taken into account in the Conditional CAPM, that will be presented in the next section.

The CAPM further takes the amount an investor wants to invest in the assets as exogenously given. However the amount to invest is a decision whether to consume today or to consume in the future. The influence of consumption is incorporated into the Consumption-based CAPM.

Finally the CAPM explains the expected returns only by a single variable, the risk of an asset relative to the market. It is reasonable to assume that other factors may as well influence the expected returns. We will therefore discuss the Arbitrage Pricing Theory as an alternative to the CAPM framework in chapter 4.6. The Intertemporal CAPM also allows for more risk factors.

Another assumption remains critical for the CAPM: the assumption that all assets are marketable. Some investment restrictions due to legislation in foreign countries are taken into account by the International CAPM, but assets such as human capital are not marketable. Therefore the market portfolio cannot be determined correctly. Roll (1977) showed that the determination of a correct market portfolio is important to achieve correct results. Only small deviations from the true market portfolio can bias the results significantly. Other investigations showed that small deviations do not have a large impact on the results.[13] It remains an open question whether the assumption of all assets to be marketable is restrictive or not.

As a results of these critiques the CAPM has been modified and today a wide variety of extensions exist. Alternatives that use a different approach to asset pricing are not frequently found and had, with the exception of the Arbitrage Pricing Theory, no great impact as well in applications as the academic literature. The remaining sections of this chapter will give an overview of the extensions of the CAPM as well as alternative approaches.

[13] See Campbell et al. (1997, pp. 213 ff.).

4. The Conditional Capital Asset Pricing Model

As we saw in the last section the performance of the traditional CAPM is only poor. It has been proposed that the reason may be the static nature of the model and the therewith following assumptions of a fixed beta and fixed risk premia. Allowing the beta to vary over time can be justified by the reasonable assumption that the relative risk and the expected excess returns of an asset may vary with the business cycle.[1]

It therefore is reasonable to use all available information on the business cycle and other relevant variables to form expectations, i.e. to use conditional moments. This gave rise to the *Conditional Capital Asset Pricing Model.*

4.1 Derivation of the model[2]

We assume that the CAPM as derived in the last section remains valid if we use conditional moments instead of unconditional moments:

$$E\left[R_i^t|\Omega_{t-1}\right] - E\left[r^t|\Omega_{t-1}\right] = \left(E\left[R_M^t|\Omega_{t-1}\right] - E\left[r^t|\Omega_{t-1}\right]\right)\beta_i^{t-1}, \tag{4.1}$$

where

$$\beta_i^{t-1} = \frac{Cov\left[R_i^t, R_M^t|\Omega_{t-1}\right]}{Var\left[R_M^t|\Omega_{t-1}\right]}. \tag{4.2}$$

R_i^t denotes the return of asset i at time t, r^t the risk free rate of return at time t, R_M^t the return of the market portfolio at time t and Ω_{t-1} the information

[1] See JAGANNATHAN/WANG (1996, pp. 4 f.).
[2] This derivation follows JAGANNATHAN/WANG (1996).

available at time $t-1$. For notational simplicity we define

$$\gamma_0^{t-1} \equiv E\left[r^t|\Omega_{t-1}\right], \tag{4.3}$$

$$\gamma_1^{t-1} \equiv E\left[R_M^t|\Omega_{t-1}\right] - E\left[r^t|\Omega_{t-1}\right] = E\left[R_M^t|\Omega_{t-1}\right] - \gamma_0^{t-1} \tag{4.4}$$

as the expected risk free rate and the expected risk premium, respectively. We can rewrite (4.1) as

$$E\left[R_i^t|\Omega_{t-1}\right] = \gamma_0^{t-1} + \gamma_1^{t-1}\beta_i^{t-1}. \tag{4.5}$$

We want to explain unconditional returns because expectations cannot be observed. Taking expectations of (4.5) we get with the law of iterated expectations:

$$E\left[R_i^t\right] = E\left[\gamma_0^{t-1}\right] + E\left[\gamma_1^{t-1}\right]E\left[\beta_i^{t-1}\right] + Cov\left[\gamma_1^{t-1}, \beta_i^{t-1}\right] \tag{4.6}$$

The risk premium, γ_1^{t-1}, and the risk, β_i^{t-1}, will in general be correlated. To see this imagine a recession in which future prospects of a company are very uncertain, then the beta will be relatively high as other companies, like consumer goods or utilities that form a part of the market portfolio, are affected much less. The risk premium will also be high to compensate the owners of the asset for the higher risk in a recession as the asset forms part of the market portfolio. In a boom relations are likely to change. Hence risk premium and risk will be correlated.

Let the sensitivity of the conditional beta to a change in the market risk premium be denoted by ϑ_i:

$$\vartheta_i \equiv \frac{Cov\left[\beta_i^{t-1}, \gamma_1^{t-1}\right]}{Var\left[\gamma_1^{t-1}\right]}. \tag{4.7}$$

We further define a residual beta as:

$$\eta_i^{t-1} = \beta_i^{t-1} - E\left[\beta_i^{t-1}\right] - \vartheta_i\left(\gamma_1^{t-1} - E\left[\gamma_1^{t-1}\right]\right). \tag{4.8}$$

The first two terms represent the difference of the conditional and the unconditional beta. The final term adjusts this difference for the deviation of the risk premium from its unconditional value. We find that

$$(4.9) \qquad \begin{aligned} E\left[\eta_i^{t-1}\right] &= E\left[\beta_i^{t-1}\right] - E\left[\beta_i^{t-1}\right] - \vartheta_i\left(E\left[\gamma_1^{t-1}\right] - E\left[\gamma_1^{t-1}\right]\right) \\ &= 0, \end{aligned}$$

$$(4.10) \qquad \begin{aligned} E\left[\eta_i^{t-1}\gamma_1^{t-1}\right] &= E\left[\beta_i^{t-1}\gamma_1^{t-1}\right] - E\left[E\left[\beta_i^{t-1}\right]\gamma_1^{t-1}\right] \\ &\quad -E\left[\vartheta_i\gamma_1^{t-1}\left(\gamma_1^{t-1} - E\left[\gamma_1^{t-1}\right]\right)\right] \\ &= E\left[\beta_i^{t-1}\right]E\left[\gamma_1^{t-1}\right] + Cov\left[\beta_i^{t-1}, \gamma_1^{t-1}\right] \\ &\quad -E\left[\beta_i^{t-1}\right]E\left[\gamma_1^{t-1}\right] \\ &\quad -\vartheta_i\left(E\left[\left(\gamma_1^{t-1}\right)^2\right] - E\left[\gamma_1^{t-1}\right]^2\right) \\ &= Cov\left[\beta_i^{t-1}, \gamma_1^{t-1}\right] - \vartheta_i Var\left[\gamma_1^{t-1}\right] \\ &= Cov\left[\beta_i^{t-1}, \gamma_1^{t-1}\right] - Cov\left[\beta_i^{t-1}, \gamma_1^{t-1}\right] \\ &= 0. \end{aligned}$$

We can now solve (4.8) for the conditional beta:

$$(4.11) \qquad \beta_i^{t-1} = E\left[\beta_i^{t-1}\right] + \vartheta_i\left(\gamma_1^{t-1} - E\left[\gamma_1^{t-1}\right]\right) + \eta_i^{t-1}.$$

We further can solve (4.7) for $Cov\left[\beta_i^{t-1}, \gamma_1^{t-1}\right]$ and insert the expression into (4.6) to obtain:

$$(4.12) \qquad E\left[R_i^t\right] = E\left[\gamma_0^{t-1}\right] + E\left[\gamma_1^{t-1}\right]E\left[\beta_i^{t-1}\right] + Var\left[\gamma_1^{t-1}\right]\vartheta_i.$$

The excess returns turn out to be a linear function of the expected beta and the sensitivity of the beta to a change in the risk premium. We want to express this relation with using the unconditional beta.

Equation (4.5) also remains valid for the market portfolio, so that we get with $\beta_M^{t-1} = 1$:

$$(4.13) \qquad E\left[R_M^t|\Omega_{t-1}\right] = \gamma_0^{t-1} + \gamma_1^{t-1},$$

and from the definition of γ_0^{t-1}:

$$\gamma_1^{t-1} = E\left[R_M^t - \gamma_0^{t-1} | \Omega_{t-1}\right]. \tag{4.14}$$

We define ε_i^t as the residual from relation (4.1):

$$\varepsilon_i^t \equiv R_i^t - \gamma_0^{t-1} - \left(R_M^t - \gamma_0^{t-1}\right)\beta_i^{t-1} \tag{4.15}$$

with

$$\begin{aligned} E\left[\varepsilon_i^t | \Omega_{t-1}\right] &= E\left[R_i^t | \Omega_{t-1}\right] - E\left[\gamma_0^{t-1} | \Omega_{t-1}\right] \\ &\quad - E\left[(R_M^t - \gamma_0^{t-1})\beta_i^{t-1} | \Omega_{t-1}\right] \\ &= \gamma_0^{t-1} + \gamma_1^{t-1}\beta_i^{t-1} - \gamma_0^{t-1} \\ &\quad - E\left[R_M^t - \gamma_0^{t-1}\right]\beta_i^{t-1} \\ &= \gamma_1^{t-1}\beta_i^{t-1} - \gamma_1^{t-1}\beta_i^{t-1} \\ &= 0, \end{aligned} \tag{4.16}$$

$$\begin{aligned} E\left[\varepsilon_i^t R_M^t | \Omega_{t-1}\right] &= E\left[R_i^t R_M^t | \Omega_{t-1}\right] \\ &\quad - E\left[\left(\gamma_0^{t-1} + \left(R_M^t - \gamma_0^{t-1}\right)\beta_i^{t-1}\right) R_M^t | \Omega_{t-1}\right] \\ &= E\left[R_i^t R_M^t | \Omega_{t-1}\right] \\ &\quad - E\left[E\left[r_i^t | \Omega_{t-1}\right] R_M^t | \Omega_{t-1}\right] \\ &= 0. \end{aligned} \tag{4.17}$$

By the law of iterated expectations we get:

$$E\left[\varepsilon_i^t\right] = E\left[E\left[\varepsilon_i^t | \Omega_{t-1}\right]\right] = 0, \tag{4.18}$$

$$E\left[\varepsilon_i^t R_M^t\right] = E\left[E\left[\varepsilon_i^t R_M^t | \Omega_{t-1}\right]\right] = 0, \tag{4.19}$$

$$\begin{aligned} E\left[\varepsilon_i^t \gamma_1^{t-1}\right] &= E\left[\varepsilon_i^t E\left[R_M^t - \gamma_0^{t-1} | \Omega_{t-1}\right]\right] \\ &= E\left[E\left[\varepsilon_i^t E\left[R_M^t - \gamma_0^{t-1} | \Omega_{t-1}\right] | \Omega_{t-1}\right]\right] \\ &= E\left[E\left[\varepsilon_i^t | \Omega_{t-1}\right] E\left[R_M^t - \gamma_0^{t-1} | \Omega_{t-1}\right]\right] \\ &= 0. \end{aligned} \tag{4.20}$$

By inserting (4.11) into (4.15) and solving for R_i^t we get

$$R_i^t = \gamma_0^{t-1} + \left(R_M^t - \gamma_0^{t-1}\right) E\left[\beta_i^{t-1}\right] + \left(R_M^t - \gamma_0^{t-1}\right)\left(\gamma_1^{t-1} - E\left[\gamma_1^{t-1}\right]\right)\vartheta_i + \left(R_M^t - \gamma_0^{t-1}\right)\eta_i^{t-1} + \varepsilon_i^t. \quad (4.21)$$

From the definition and the properties of the covariance we get with (4.21):

$$Cov\left[R_i^t, R_M^t\right] = Cov\left[\gamma_0^{t-1}, R_M^t\right] + Cov\left[R_M^t - \gamma_0^{t-1}, R_M^t\right] E\left[\beta_i^{t-1}\right] + Cov\left[\left(R_M^t - \gamma_0^{t-1}\right)\left(\gamma_1^{t-1} - E\left[\gamma_1^{t-1}\right]\right)\vartheta_i, R_M^t\right] + Cov\left[\left(R_M^t - \gamma_0^{t-1}\right)\eta_i^{t-1}, R_M^t\right], \quad (4.22)$$

$$Cov\left[R_i^t, \gamma_i^{t-1}\right] = Cov\left[\gamma_0^{t-1}, \gamma_i^{t-1}\right] + Cov\left[R_M^t - \gamma_0^{t-1}, \gamma_i^{t-1}\right] E\left[\beta_i^{t-1}\right] + Cov\left[\left(R_M^t - \gamma_0^{t-1}\right)\left(\gamma_1^{t-1} - E\left[\gamma_1^{t-1}\right]\right)\vartheta_i, \gamma_i^{t-1}\right] + Cov\left[\left(R_M^t - \gamma_0^{t-1}\right)\eta_i^{t-1}, \gamma_i^{t-1}\right]. \quad (4.23)$$

It has been shown by JAGANNATHAN/WANG (1996, pp. 38 ff.) that the last term in (4.22) and (4.23) becomes zero if we assume the residual betas, η_i^{t-1}, to be uncorrelated with market conditions.

We define

$$\beta_i = \frac{Cov\left[R_i^t, R_M^t\right]}{Var\left[R_M^t\right]} \quad (4.24)$$

as the traditional market beta which measures the unconditional risk and

$$\beta_i^\gamma = \frac{Cov\left[R_i^t, \gamma_1^{t-1}\right]}{Var\left[\gamma_1^{t-1}\right]} \quad (4.25)$$

as the premium beta which measures the risk from a varying beta. We can substitute (4.24) and (4.25) into (4.22) and (4.23), respectively, and obtain after solving for β_i and β_i^γ:

$$\begin{aligned}(4.26)\qquad \beta_i &= \frac{Cov\left[\gamma_0^{t-1}, R_M^t\right]}{Var\left[R_M^t\right]} \\ &+\frac{Cov\left[R_M^t-\gamma_0^{t-1}, R_M^t\right]}{Var\left[R_M^t\right]}E\left[\beta_i^{t-1}\right] \\ &+\frac{Cov\left[\left(R_M^t-\gamma_0^{t-1}\right)\left(\gamma_1^{t-1}-E\left[\gamma_1^{t-1}\right]\right), R_M^t\right]}{Var\left[R_M^t\right]}\vartheta_i,\end{aligned}$$

$$\begin{aligned}(4.27)\qquad \beta_i &= \frac{Cov\left[\gamma_0^{t-1}, \gamma_1^{t-1}\right]}{Var\left[\gamma_1^{t-1}\right]} \\ &+\frac{Cov\left[R_M^t-\gamma_0^{t-1}, \gamma_1^{t-1}\right]}{Var\left[\gamma_1^{t-1}\right]}E\left[\beta_i^{t-1}\right] \\ &+\frac{Cov\left[\left(R_M^t-\gamma_0^{t-1}\right)\left(\gamma_1^{t-1}-E\left[\gamma_1^{t-1}\right]\right), \gamma_1^{t-1}\right]}{Var\left[\gamma_1^{t-1}\right]}\vartheta_i.\end{aligned}$$

If we define $b_{10} \equiv \frac{Cov\left[\gamma_0^{t-1},R_M^t\right]}{Var\left[R_M^t\right]}$, $b_{11} \equiv \frac{Cov\left[R_M^t-\gamma_0^{t-1},R_M^t\right]}{Var\left[R_M^t\right]}$, $b_{12} \equiv \frac{Cov\left[\left(R_M^t-\gamma_0^{t-1}\right)\left(\gamma_1^{t-1}-E\left[\gamma_1^{t-1}\right]\right),R_M^t\right]}{Var\left[R_M^t\right]}$, $b_{20} \equiv \frac{Cov\left[\gamma_0^{t-1},\gamma_1^{t-1}\right]}{Var\left[\gamma_1^{t-1}\right]}$, $b_{21} \equiv \frac{Cov\left[R_M^t-\gamma_0^{t-1},\gamma_1^{t-1}\right]}{Var\left[\gamma_1^{t-1}\right]}$, $b_{22} \equiv \frac{Cov\left[\left(R_M^t-\gamma_0^{t-1}\right)\left(\gamma_1^{t-1}-E\left[\gamma_1^{t-1}\right]\right),\gamma_1^{t-1}\right]}{Var\left[\gamma_1^{t-1}\right]}$, $b = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix}$ and $B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$ we can write (4.26) and (4.27) in vector form as

$$(4.28)\qquad \begin{bmatrix} \beta_i \\ \beta_i^\gamma \end{bmatrix} = b + B \begin{bmatrix} E\left[\beta_i^{t-1}\right] \\ \vartheta_i \end{bmatrix}.$$

If $E\left[\beta_i^{t-1}\right]$ and ϑ_i are linear dependent, i.e. $E\left[\beta_i^{t-1}\right] = d_0 + d_1\vartheta_i$, we get from (4.28):

$$\begin{aligned}(4.29)\qquad \beta_i &= b_{10} + b_{11}(d_0 + d_1\vartheta_i) + b_{12}\vartheta_i \\ &= (b_{10} + b_{11}d_0) + (b_{11}d_0 + b_{12})\vartheta_i.\end{aligned}$$

By inserting (4.29) into (4.12) we get

$$
\begin{aligned}
(4.30) \qquad E\left[R_i^t\right] &= E\left[\gamma_0^{t-1}\right] + E\left[\gamma_1^{t-1}\right](d_0 + d_1\vartheta_i) + Var\left[\gamma_1^{t-1}\right]\vartheta_i \\
&= E\left[\gamma_0^{t-1}\right] + E\left[\gamma_1^{t-1}\right]d_0 \\
&\quad + \left(E\left[\gamma_1^{t-1}\right]d_1 + Var\left[\gamma_1^{t-1}\right]\right)\vartheta_i \\
&= E\left[\gamma_0^{t-1}\right] + E\left[\gamma_1^{t-1}\right]d_0 \\
&\quad - \frac{\left(E\left[\gamma_1^{t-1}\right]d_1 + Var\left[\gamma_1^{t-1}\right]\right)(b_{10} + b_{11}d_0)}{b_{12} + b_{11}d_0} \\
&\quad + \frac{E\left[\gamma_1^{t-1}\right]d_1 + Var\left[\gamma_1^{t-1}\right]}{b_{12} + b_{11}d_0}\beta_i \\
&= a_0 + a_1\beta_i
\end{aligned}
$$

with $a_0 = E\left[\gamma_0^{t-1}\right] + E\left[\gamma_1^{t-1}\right]d_0 - \frac{\left(E\left[\gamma_1^{t-1}\right]d_1 + Var\left[\gamma_1^{t-1}\right]\right)(b_{10}+b_{11}d_0)}{b_{12}+b_{11}d_0}$ and $a_1 = \frac{E\left[\gamma_1^{t-1}\right]d_1+Var\left[\gamma_1^{t-1}\right]}{b_{12}+b_{11}d_0}$. In this case the conditional CAPM has the same form as the unconditional form presented in the last section. But if $E\left[\beta_i^{t-1}\right]$ is not a linear function of ϑ_i we are able to invert equation (4.28):

$$
(4.31) \qquad \begin{bmatrix} E\left[\beta_i^{t-1}\right] \\ \vartheta_i \end{bmatrix} = B^{-1}b - B^{-1}\begin{bmatrix} \beta_i \\ \beta_i^{\gamma} \end{bmatrix}.
$$

By rewriting (4.12) we get in vector form:

$$
(4.32) \qquad E\left[R_i^t\right] = E\left[\gamma_0^{t-1}\right] + \left[E\left[\gamma_1^{t-1}\right] Var\left[\gamma_1^{t-1}\right]\right]\begin{bmatrix} E\left[\beta_i^{t-1}\right] \\ \vartheta_i \end{bmatrix}.
$$

Inserting (4.31) this becomes

$$
\begin{aligned}
(4.33) \qquad E\left[R_i^t\right] &= E\left[\gamma_0^{t-1}\right] + \left[E\left[\gamma_1^{t-1}\right] Var\left[\gamma_1^{t-1}\right]\right]B^{-1}b \\
&\quad - \left[E\left[\gamma_1^{t-1}\right] Var\left[\gamma_1^{t-1}\right]\right]B^{-1}\begin{bmatrix} \beta_i \\ \beta_i^{\gamma} \end{bmatrix}.
\end{aligned}
$$

With $B^{-1} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$, $e_0 = E\left[\gamma_0^{t-1}\right] + \left[E\left[\gamma_1^{t-1}\right], Var\left[\gamma_1^{t-1}\right]\right]B^{-1}b$, $e_1 = E\left[\gamma_1^{t-1}\right]c_{11} + Var\left[\gamma_1^{t-1}\right]c_{21}$ and $e_2 = E\left[\gamma_1^{t-1}\right]c_{12} + Var\left[\gamma_1^{t-1}\right]c_{22}$we can rewrite (4.33) as

$$
(4.34) \qquad E\left[R_i^t\right] = e_0 + e_1\beta_i + e_2\beta_i^{\gamma}.
$$

The expected return depends linearly on the market risk and the risk of a change in the market risk, i.e. it depends on two different beta. Although we have the dependence on two beta, this model is not a special case of other models that will be discussed later having multi-beta structures. The original form had a single (conditional) beta, only to derive the unconditional form this second beta turned out, no other source of risk than the market risk has been added. This second beta is due to the unobservability of expectations. Any risk factor that we can determine can change over time, i.e. we could derive the conditional version for various sources of risk, when determining the unconditional form for every risk factor such a second beta would turn out, hence the Conditional CAPM is a generalization of the unconditional form and not a generalization to include other risk factors.

4.2 Empirical results

Models with time varying betas and risk premia have attracted increased attention in recent years.[3] The reason on one hand is the empirical evidence that covariances, variances and risk premia are not constant over time. On the other hand gives the poor performance of the traditional CAPM rise to modifications of this model.

By introducing time varying betas and risk premia the conditional CAPM remains to be a static model from its nature. Although a dynamic model would capture reality more appropriate, the conditional CAPM performs much better than its unconditional form.[4] The model by Fama/French (1992), which adds other factors like the book-to-market ratio or firm size, does not perform approximately well and including other factors into the conditional CAPM does not improve the

[3] See e.g. Ferson (1995), Ferson/Harvey (1991), Ferson/Harvey (1996), Jagannathan/Wang (1996) and their references.

[4] See Jagannathan/Wang (1996, pp. 18 ff.) and Ferson/Harvey (1996, pp. 9 ff.). It turns out that it even outperforms several multi-beta and dynamic models to be presented in the following sections.

results much, suggesting that their factors are of no real importance.[5]

Jagannathan/Wang (1996) also showed that by including other risk factors, different from those of Fama/French (1992), the performance can be increased significantly by using monthly instead of yearly data. This gives rise to other models allowing for more general risk factors and dynamic models.

Summarizing can be said that the Conditional CAPM fits the data much better than the traditional CAPM, but especially for explaining asset prices in the short run, it fails.

[5] See Jagannathan/Wang (1996, pp. 31 ff.).

5. Models of Changing Volatility of Asset Returns

The models considered thus far were all based on an economic theory assuming rational investors optimizing an objective function. Despite the variety of models that will be presented in the remaining sections, there exist a number of of effects in asset markets that cannot be explained by any of these theories. One such effect is the observation that the volatility, i.e. variance, of asset returns change over time. This effect can be found in the long run using monthly or quarterly returns, as well as in the short run, using intra-day returns (high frequency data).

While we could assign this effect in the long run to the effect of rational speculation by stating that over time the dispersion of fundamental values between investors changes,[1] to make such an assumption in the short-run for high frequency data is not very reasonable.

The lack of a theory explaining this effect and the strong demand from investors to get tools for basing their decisions on, gave rise to a class of econometric models introduced by Engle (1982) and Bollerslev (1986). Their models of conditional heteroskedasticity, together with a large number of variations developed on their basic work, have become widely applied to model the returns of assets, although these models have no economic theory that support them.

[1] See chapter 4.11.2.

5.1 The ARCH-models

The first to propose a model of changing variances was Engle (1982) with his *Autoregressive Conditional Heteroskedasticity* (*ARCH*) model. It assumes asset returns to be given by a linear combination of explanatory variables x_t and some parameters β and an error term ε_t, that has an expected value of zero given all available information. The information (explanatory variables) is restricted to past returns and error terms:

$$r_t = x_t'\beta + \varepsilon_t, \tag{5.1}$$

$$\begin{aligned} E\left[\varepsilon_t|\Omega_{t-1}\right] &= 0, \\ E\left[\varepsilon_t\varepsilon_{t+i}|\Omega_{t-1}\right] &= 0, \qquad i \neq 0, \\ Var\left[\varepsilon_t|\Omega_{t-1}\right] &= E\left[\varepsilon_t^2|\Omega_{t-1}\right] = h_t. \end{aligned}$$

The variance of the error term is now assumed to vary aver time by proposing it to follow a qth order autoregressive process, i.e. an AR(q):

$$\begin{aligned} h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2, \\ \alpha_j &\geq 0 \quad (j = 0, \ldots, q). \end{aligned} \tag{5.2}$$

This specification was defined as the ARCH(q)-process.

An extension to the *Generalized Autoregressive Conditional Heteroskedasticity* (*GARCH*) model was provided by Bollerslev (1986). He proposes the conditional variance to follow a (p,q) Autoregressive Moving Average process, ARMA(p,q). To form the GARCH(p,q)-process equation (5.2) has to be replaced by

$$\begin{aligned} h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \gamma_i h_{t-i}, \\ \alpha_j &\geq 0 \quad (j = 1, \ldots, q), \\ \gamma_j &\geq 0 \quad (j = 1, \ldots, p). \end{aligned} \tag{5.3}$$

As such models are difficult to estimate from the data, most applications restrict themselves to a GARCH(1,1)-process:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma h_{t-1}. \tag{5.4}$$

As nearly all asset pricing theories use the variance of the return as a measure for the risk of an asset and these theories agree that a higher risk has to be compensated by a higher expected return, it is reasonable to include a term for the variance in the expression of the return of equation (5.1). These considerations gave rise to the *ARCH-in-Mean* (*ARCH-M*) introduced by ENGLE ET AL. (1987). They replaced equation (5.1) by

$$r_t = x_t' \beta + \delta h_t + \varepsilon_t, \tag{5.5}$$

where the conditional variance of ε_t is given by (5.2). This model can easily be genaralized to a *GARCH-in-Mean* (*GRARCH-M*) by using (5.3) as the process governing conditional variances.

When investigating not only a single, but a larger number of assets, it is useful to assume that all assets are governed by the same process. Therefore ENGLE/KRONER (1995) provided a multivariate version of the GARCH-process. If we let B denote the matrix of parameters x_t, ε_t the vector of error terms and r_t the vector of returns, we get:

$$r_t = Bx_t + \varepsilon_t, \tag{5.6}$$

$$\begin{aligned} E\left[\varepsilon_t | \Omega_{t-1}\right] &= 0, \\ E\left[\varepsilon_t' \varepsilon_{t+i} | \Omega_{t-1}\right] &= 0, \qquad i \neq 0, \\ Var\left[\varepsilon_t | \Omega_{t-1}\right] &= E\left[\varepsilon_t \varepsilon_t' | \Omega_{t-1}\right] = H_t, \end{aligned}$$

$$H_t = K + \sum_{i=1}^{p} A_i H_{t-i} A_i' + \sum_{i=1}^{q} c_i \varepsilon_{t-i} \varepsilon_{t-i}' C_i'. \tag{5.7}$$

The multivariate GARCH-M can easily be obtained by replacing (5.6) with

$$r_t = Bx_t + Dh_t + \varepsilon_t. \tag{5.8}$$

where $h_t = vech(H_t)$. These multivariate forms allow not only the variances of asset returns to vary, but also the covariances between asset returns. Therewith these models are close to that what can be observed from real asset returns, although the specification of the model may not be appropriate.

Beside the models presented here a wide variety of other specifications exists in the literature.[2] A variation that is frequently found in empirical investigations is the *Exponential GARCH(EGARCH)* developed by NELSON (1991). In empirical investigations it often turns out that this model fits the data best and is easier to estimate. The EGARCH assumes the following form of the conditional variance:

$$h_t = \exp\left[\alpha_0 + \sum_{j=1}^{\infty} \alpha_j \left(|v_{t-j}| - E\left[v_{t-j}\right] + \psi v_{t-j}\right)\right], \tag{5.9}$$

where $v_{t-j} = \frac{\varepsilon_{t-j}}{\sqrt{h_{t-j}}}$. The EGARCH specification allows with $\psi < 0$ to model asymmetric shocks to volatility. If $-1 < \psi < 0$ a positive shock, i.e. $\varepsilon_{t-j} > 0$, increases volatility less than a negative shock ($\varepsilon_{t-j} < 0$) in subsequent periods. For $\psi < 1$ a positive shock reduces the future variances and a negative shock increases it.[3]

5.2 The relation to asset pricing models

A large number of empirical investigations using some forms of ARCH-processes are reported in the literature, most investigations use models based on GARCH(1,1) for computational simplicity. They show a strong support that asset returns follow this specification, often the strongest evidence is found for a EGARCH(1,1)-process.[4]

The conditional CAPM introduced in section 4.4 allowed for time varying vari-

[2] Overviews of these models can be found in BOLLERLEV ET AL. (1992) or HAMILTON (1994, pp. 665 ff.).

[3] See HAMILTON (1994, p. 668).

[4] See BOLLERLEV ET AL. (1992) for an overview. DUNNE (1997) uses a multivariate GARCH-M to model the time varying risk-premium using the FAMA/FRENCH (1993) and FAMA/FRENCH (1995) factors and finds strong evidence of this model.

ances and covariances by taking their conditional values. As the only explanatory variable the market risk has been used, but it has already been stated that that other explanatory variables could easily be incorporated.

The ARCH-models presented here shows a simple specification of forming conditional expectations on variances and covariances. It allows only past variances, covariances and error terms to form the information set and combines them in a linear form to form new expectations.

We should therefore interpret the ARCH-models as a specification of the conditional asset pricing model. So we should not be surprised that ARCH-models show strong support of their specification as it falls into the class of conditional models and the Conditional CAPM is also strongly supported by investigations not using ARCH specifications.

The ARCH-models fit the data very well, as well in the short run as in the long run, especially the variances are well captured. To find an economic model how to justify a simple ARCH framework remains an unsolved problem up to now.

6. The Arbitrage Pricing Theory

Beside the problem of identifying the market portfolio and the critiques concerning the mean-variance criterion, a critical point in the concept of the CAPM is the aggregation of all risks into a single risk factor, the market risk. This aggregation is useful for optimal or at least well diversified portfolios, but for the explanation of returns of individual assets this aggregation may be problematic. It is well observable that assets are not only driven by general factors like the market movement, but that industry or country specific influences also have a large impact on returns.

This section presents an alternative to the CAPM, the *Arbitrage Pricing Theory* (*APT*) as first introduced by ROSS (1976).[1]

6.1 Derivation of the APT

Assume that the returns are generated by the following linear structure:

$$\tilde{R}_i = \mu_i + \beta_i \tilde{\delta} + \tilde{\varepsilon}_i, \tag{6.1}$$

where $\tilde{R}_i$ denotes the realized return of asset i, μ_i the unconditional expected return, $\tilde{\delta}$ a vector of different risk factors, β_i a vector representing the influence each risk factor has on the asset return and $\tilde{\varepsilon}_i$ an error term summarizing the

[1] ROSS (1976) analyzes a sequence of economies over time and shows the convergence of the returns to his theory. Without changing the arguments we will use different assets instead of different points of time and show that the relationship holds if the number of assets increases to infinity. This change in argumentation is widely done in the literature, see SCHNELLER (1990, p. 2).

effects not covered by the model.[2] We make the following assumptions on this structure:

$$\begin{aligned} E\left[\tilde{\delta}\right] &= 0, \\ E\left[\tilde{\varepsilon}_i\right] &= 0. \end{aligned} \tag{6.2}$$

These assumptions state that the influence of effects not covered have on average no influence on the returns, i.e. there is no systematic bias. The factors having an influence on the returns are assumed to be normalized, i.e. only deviations from their average values are considered, the effect of the level of these factors are summarized in μ_i.

we do not have to assume that the ε_i are independent of each other, it is sufficient if they are not too dependent on each other, such that the law of large number can be applied. We need for all i,j $(i \neq j)$ that[3]

$$E\left[\tilde{\varepsilon}_i \tilde{\varepsilon}_j\right] = Cov\left[\tilde{\varepsilon}_i, \tilde{\varepsilon}_j\right] \approx 0. \tag{6.3}$$

By the law of large numbers we find a vector $x_n = (x_1, \ldots, x_n)'$ such that $\lim_{n\to\infty} x_n' \tilde{\varepsilon} = 0$ where $\tilde{\varepsilon} = (\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$, i.e. in a well diversified portfolio the error terms have no influence on the return of a portfolio. We therefore interpret x_i as the weight asset i has in such a portfolio.

In general for any portfolio x we have

$$\begin{aligned} x'\tilde{R} &= x'\mu + x'\beta\tilde{\delta} + x'\tilde{\varepsilon} \\ &\approx x'\mu + x'\beta\tilde{\delta}, \end{aligned} \tag{6.4}$$

where $\tilde{R} = \left(\tilde{R}_1, \ldots, \tilde{R}_n\right)$, $\mu = (\mu_1, \ldots, \mu_n)$ and $\beta = (\beta_1, \ldots, \beta_n)$. The term $x'\beta$ denotes the risk from the common factors that cannot be eliminated by diversification, i.e. the systematic risk, while $x'\tilde{\varepsilon}$ can be eliminated by diversification, i.e. it is the unsystematic risk.

[2] It should be noted that this structure with the assumptions made below is identical to the problem stated in factor analysis.

[3] See Ross (1976, p. 342).

An *arbitrage portfolio* is defined as a portfolio with no risk, no net investment, but a positive certain return.[4] Hence in an arbitrage portfolio neither systematic nor unsystematic risk must be present and the following restrictions apply:

$$
\begin{aligned}
x'\tilde{\varepsilon} &= 0, \\
x'\beta &= 0, \\
x'\iota &= 0,
\end{aligned}
\tag{6.5}
$$

where $\iota = (1, \ldots, 1)'$. Inserting these conditions into 6.4 gives us for an arbitrage portfolio:

$$x'\tilde{R} = x'\mu > 0. \tag{6.6}$$

In equilibrium we must assure that no arbitrage is possible, i.e.

$$x'\mu = 0.^5 \tag{6.7}$$

If we continue to apply the mean-variance criterion, which is not implied by the APT but mostly done for conventional reasons, the condition of no arbitrage possibilities assures that with the introduction of more assets the efficient frontier does not converge to a vertical line in the (μ,σ)-plane, i.e. $\frac{\partial \mu}{\partial \sigma} < \infty$.[6]

The assumptions underlying the APT as presented by Schneller (1990, pp. 2 ff.) and Ross (1976, pp. 347 and 351) are summarized in table 6.1.

Let us consider an arbitrage portfolio fulfilling restrictions (6.5) and the expected return of this portfolio is

$$x'\mu = c > 0, \tag{6.8}$$

[4] See Schneller (1990, p.3).
[5] From the definition of an arbitrage portfolio it would be sufficient to state $x'\mu \leq 0$, but if $x'\mu < 0$ we could change the signs of every element in x and get with this new portfolio $x'\mu > 0$ and arbitrage would be possible. Hence we have to state the condition that $x'\mu = 0$.
[6] See Schneller (1990, p.17).
[8] This assumption has been shown by Ross (1976, p. 352) to be possible to relax in that way that at least one individual has to believe the return to follow (6.1) and this individual must have a considerable part invested into each asset, i.e. his share in holding the asset does not converge to zero if more assets are added.
[8] This implies also the possibility to borrow and lend the risk free asset, if such an asset exists, at the same rate.

- The returns are assumed by investors to follow equation (6.1)[7]
- Investors are risk averse with a finite Arrow-Pratt measure of risk aversion
- No transaction costs or taxes exist
- No restrictions on short sales for any asset[8]
- In equilibrium no arbitrage possibilities exist
- For at least one asset the possible loss from holding the asset is limited to $t < \infty$
- Every asset wants to be held by investors, i.e. the total demand for every asset is positive
- Homogeneity of beliefs, i.e. all investors expect the same $\mu_i < \infty$ and agree on β_i

Table 6.1: Assumptions of the APT

where

$$c = m + t \tag{6.9}$$

with $m = x_0\mu$ denoting the expected return of the optimal portfolio of the investor. From the limitation that $\mu_i < \infty$ for all assets we know that $m < \infty$. We will therefore consider an arbitrage portfolio whose certain return is larger than the expected return of the optimal portfolio plus the maximum loss associated to the asset with limited liability, i.e. the arbitrage portfolio should be preferred to all other portfolios. We do not assume this portfolio to have no unsystematic risk as it is not necessarily well diversified, we only assume it has no systematic risk.

Before deriving the optimal arbitrage portfolio we have to state some preliminaries from probability theory. Ross (1976, p. 359) states that if $\tilde{y}_n \to a$ in quadratic mean, then $E\left[U\left(\tilde{y}_n\right)\right] \to U(a)$ in quadratic mean.

Suppose that the unsystematic risk $x'Vx$, V denoting the covariance matrix, of the arbitrage portfolio converges to zero in quadratic mean, i.e. $x'Vx \to$

0. We then find that with the above result $E\left[U\left(x'\tilde{R}-t\right)\right] \to U(x'\mu - t) = U(c-t) > U(m)$. By the ε-criterion of convergence there exists a number of assets n such that $E\left[U\left(x_n'\tilde{R}-t\right)\right] > U(m_n)$, hence the optimal portfolio x_{0n} with its return m_n would no longer be optimal. Hence $x'Vx$ does not converge to zero and there exists a number $a > 0$ such that

$$x'Vx \geq a. \tag{6.10}$$

We now can determine the optimal arbitrage portfolio. From the theory of portfolio selection we know that the only efficient portfolio for a given expected return is the portfolio which has the lowest variance (risk). The risk of the arbitrage portfolio is only the unsystematic risk $x'Vx$. We have to apply the restrictions

$$\begin{aligned} x'\mu &= 0, \\ x'\beta &= 0, \\ x'\iota &= 0. \end{aligned}$$

By defining $W = (\mu, \beta, \iota)'$ and $g = (c, 0, 0)$ we can rewrite these restrictions as

$$x'W = g. \tag{6.11}$$

The Lagrange function shows up to be

$$L = x'Vx - 2\lambda(x'W - g). \tag{6.12}$$

The first order conditions for a minimum are

$$\begin{aligned} \frac{\partial L}{\partial x} &= 2Vx - 2\lambda W = 0, \\ \frac{\partial L}{\partial \lambda} &= -2(x'W - g) = 0, \end{aligned} \tag{6.13}$$

which solves to[9]

$$Vx = \lambda W, \tag{6.14}$$

[9] The second derivative with respect to x is $2V$. As V is a covariance it is positive definite and the second order condition for a minimum is fulfilled.

$$x'W = g. \tag{6.15}$$

Solving (6.14) for x and inserting into (6.15) gives[10]

$$\left(W'V^{-1}W\right)\lambda = g. \tag{6.16}$$

We further find

$$x'Vx = x'W\lambda = g\lambda = g'\left[W'V^{-1}W\right]^{-1} g \geq a > 0. \tag{6.17}$$

Ross (1976, pp. 357 f.) has proofed that there exists a vector a^* and a number $A > 0$ such that with $a^* g = 1$

$$(Wa^*)'(Wa^*) \leq A < \infty. \tag{6.18}$$

If we define $ca^* = (1, -\gamma, -\rho)'$ we have $a^* g = 1$ and (6.18) becomes

$$\sum_{i=1}^{n} (\mu_i - \beta_i\gamma - \rho)^2 \leq c^2 A < \infty. \tag{6.19}$$

As (6.19) has to hold for any n, also for $n = \infty$ this implies that on average

$$(\mu_i - \beta_i\gamma - \rho)^2 \to 0. \tag{6.20}$$

A finite number of assets may not fulfill (6.20) but all of the remaining assets have to fulfill it,[11] hence nearly all assets have expected returns according to

$$\mu_i \to \beta_i\gamma + \rho, \tag{6.21}$$

what implies

$$\mu_i \approx \rho + \beta_i\gamma. \tag{6.22}$$

If we assume asset 1 to be a zero-beta portfolio, i.e. $x_1\beta_1 = 0$ then the return of this asset becomes in the absence of systematic risk:

$$\begin{aligned} x_1\tilde{R}_1 &= x_1\mu_1 + x_1\beta_1\tilde{\delta} + x_1\tilde{\varepsilon}_1 \\ &= x_1\mu_1 \\ &= x_1\beta_1\gamma + x_1\rho \\ &= x_1\rho. \end{aligned} \tag{6.23}$$

[10] We assume all matrices to have full rank for simplicity. Ross (1976) showed how to change the structure of the model to deal with matrices having not full rank.

[11] See Campbell et al. (1997, p. 221).

Hence we have found that

$$\tilde{R}_1 = \rho \equiv r. \tag{6.24}$$

ρ is the return of the zero-beta portfolio. If there is a riskless asset it is the return of this asset.

Assume now that there exist portfolios which have only a risk on one factor, l, and that this risk is a unit, i.e. $x'\beta^l = 1$ and $x'\beta^s = 0$ for all $s \neq l$. In this case (6.22) becomes with inserting (6.24) and $x'\iota = 1$ as normalization:

$$x'\mu \approx x'r\iota + x'\beta\gamma = r + \gamma^l \tag{6.25}$$

or

$$\gamma^l \approx x'\mu - r = \mu^l - r, \tag{6.26}$$

where $\mu^l \equiv x'\mu$ is the expected return of the portfolio with only a unit risk in factor l and no risk in the other factors. Therefore we can interpret γ^l as the risk premium for having one unit of risk in factor l. β_i^l then represents the amount of risk asset i has from factor l.

By inserting (6.24) and (6.26) into (6.22) we get

$$\mu_i \approx r + (\mu^* - r\iota)\beta_i, \tag{6.27}$$

where $\mu^* = (\mu^1, \dots, \mu^k)$

The exact relation only holds for an infinite number of assets, but as we find a very large number of assets the deviation will be very small.[12]

At a first glance we could interpret the APT as a generalization of the CAPM to a multibeta model. The same structure of the result in equation (6.27) as in the CAPM seems to support this view. But it has clearly to be pointed out that the models differ substantially in their assumptions. The CAPM is concerned

[12] See Ross (1976, pp. 353 f.).

to find an equilibrium of the market by holding optimal portfolios as implied by portfolio theory, whereas the APT finds this equilibrium by ruling out arbitrage possibilities. We will address this distinction between the CAPM and the APT in the next section that provides a very similar extension of the CAPM.

The APT allows to include other sources of risk than only the market risk, e.g. industry specific factors. Furthermore a market portfolio has not to be determined, consequently we also do not face the problem of excluding assets from the considerations, with a sufficient number of asset this relation should hold. Also other measures of risk than variances and covariances could be used, how to determine the betas is not predetermined by this theory.

As noted in section 4.3.2 Fama/French (1992) and Fama/French (1993) find evidence that other variables are able to explain the observed returns better than the market risk. The APT could be a framework to find a justification of their results on a sound theoretical basis. The next section will therefore give a short overview of the empirical findings concerning the APT.

6.2 Empirical evidence

The first problem to solve in applying the APT is to identify the risk factors $\tilde{\delta}$. Risk factors can either be constructed by finding a portfolio of assets that has a high correlation with a certain risk, this portfolio is called a *factor portfolio*, or by using other variables, such as macroeconomic data, e.g. GDP. The advantage of the former approach is that expected returns for the risk factor, μ^l can easily be determined from the market and the risk β_i can also be estimated from market data. The identification of these parameters for macroeconomic data imposes much more difficulties. This is the reason why in most cases factor portfolios are used.[13]

[13] See Campbell et al. (1997, pp. 222 ff.) for the statistical methods of forming factor portfolios and estimating the relevant parameters.

To identify the systematic risk and hence the characteristics of the factor portfolios we can either use theoretical considerations or statistical methods to identify these risks. Widely used statistical methods are factor analysis and principal components method.[14]

There exists a large number of surveys investigating the explanation of asset returns using APT.[15] The factors mostly identified in these studies are related to dividends or earnings, book-market relations, the size of a company and the variance of asset returns.[16] Most investigations show that three to five factors are sufficient to explain the observed returns, adding more factors does not improve the result substantially.[17]

The investigations cited give evidence that the APT can explain the observed returns quite good for long and medium time horizons. For time horizons below one year they are not able to explain the data adequately. Compared to the present value model the time horizon can be reduced significantly from four to about one year, but as in the CAPM there remain many effects that cannot be explained sufficiently.

The assumption of a linear relation between the assets in the CAPM by the covariances is replaced by the assumption of a linear relationship with risk factors. Like in the CAPM this assumption limits the theory as nonlinear assets, like derivatives, cannot be modeled adequately. The advantage of the APT in this case is that it is not necessary to form a market portfolio and to include these assets, it enables to exclude human capital or real estate. It enables also to restrict the analysis to a certain group of assets, provided that the number of assets is sufficiently large that the approximation in (6.27) holds. The more assets are included the more precise the findings should be, with restricting to only a few

[14] See Campbell et al. (1997, pp. 233 ff.) for a short overview of these methods or Brachinger/Ost (1996) for a detailed description.

[15] See the references in Fama/French (1996) for an overview of the main contributions.

[16] As we do not need a market portfolio for the APT the variance represents an appropriate risk measure instead of the covariance in the CAPM.

[17] See Fama/French (1992), Fama/French (1993) and Fama/French (1996).

assets the pricing relation does not break down as in the CAPM, it only becomes less precise, i.e. we should find more noise.

7. The Intertemporal Capital Asset Pricing Model

The models of asset pricing considered so far had one feature in common: they were all static. The amount invested into assets was fixed for a given period of time and the amounts invested into each asset could not be changed. At the end of a time period it was assumed that the investors consume their wealth.

A more realistic setting would be to allow investors to change the amounts invested into each asset and also to withdraw a part of their investment for immediate consumption. Not only the amount invested into each asset but also the fraction of the wealth invested into assets will become an endogenous variable, while in the previous models this fraction was fixed.

The *Intertemporal Capital Asset Pricing Model* (*ICAPM*) as first developed by Merton (1973) takes these considerations in account.

7.1 The model [1]

In the ICAPM we assume a perfect market, i.e. all assets have limited liability, we face no transaction costs or taxes, assets are infinitely divisible, each investor believes that his decision does not affect the market price, the market is always in equilibrium, hence we have no trades outside the equilibrium prices, investors can borrow and and lend without any restrictions all assets at the same rate.

[1] The presentation follows Merton (1973) with some proofs based on Merton (1990, chs. 4, 5, 11 and 15)

Further assumptions are that trading takes place continuously, i.e. all investors can trade at every point of time.[2] All variables that can explain the prices and price changes of the assets (the *state variables*) follow a joint Markov process.[3] The state variables are further assumed to change continuously over time, i.e. no jumps are allowed.

If we let P_t^i denote the price of asset i at time t, Ω_t the information available at time t and h the number of time units, we have for the expected return of asset i, μ_i, and the variance of the returns, σ_i^2, per unit of time:

$$\mu_i = \frac{1}{h} E\left[\frac{P_{t+h}^i - P_t^i}{P_t^i} | \Omega_t\right], \tag{7.1}$$

$$\sigma_i^2 = \frac{1}{h} E\left[\left(\frac{P_{t+h}^i - P_t^i}{P_t^i} - h\mu_i\right)^2 | \Omega_t\right]. \tag{7.2}$$

If we assume that μ_i and σ_i^2 exist and are finite and that $\lim_{h\to 0} \sigma_i^2 > 0$, i.e. by trading at every point of time (with a very short time horizon) the uncertainty cannot be eliminated.[4]

Table 7.1 summarizes the main assumptions of the ICAPM.

Define $y_i(t)$ to be iid $N(0,1)$ distributed,[5] we then can write the return dynamics implied by (7.2) with solving for $\frac{P_{t+h}^i - P_t^i}{P_t^i}$:

$$\frac{P_{t+h}^i - P_t^i}{P_t^i} = h\mu_i + \sqrt{h}\sigma_i y_i(t). \tag{7.3}$$

Taking the limit of (7.3) with respect to h we get the differential equation of the return process:

$$\frac{dP^i}{P^i} = \lim_{h\to 0} \frac{P_{t+h}^i - P_t^i}{P_t^i} = \mu_i dt + \sqrt{dt}\sigma_i y_i(t). \tag{7.4}$$

---

[2] FAMA (1996) developed a model in discrete time obtaining similar results.

[3] In a Markov process the distribution of the random variable only depends on the most recent past realization of the process. For more details on Markov processes see HAMILTON (1994, pp. 678 ff.).

[4] See MERTON (1973, pp. 868 ff.).

[5] The assumption of normality is not necessary for the results, but it allows to concentrate the analysis on the well known class of Wiener and Ito processes. See MERTON (1973, p. 873).

• All assets have limited liability
• No transaction costs and taxes
• No dividends are paid
• All assets are infinitely divisible
• All investors believe that their decisions do not influence the market price
• All trades take place in equilibrium
• Unrestricted borrowing and lending of all assets at the same conditions
• Trading takes place continuously
• Uncertainty cannot be eliminated by a continuous revision of the portfolio
• The state variables follow a joint Markov process
• The state variables change continuously

Table 7.1: Assumptions of the ICAPM

Define dz_i to be a Wiener process:

$$dz_i = y_i(t)\sqrt{dt}. \tag{7.5}$$

Inserting (7.5) into (7.4) gives us the result that returns follow an Ito process:

$$\frac{dP^i}{P^i} = \mu_i dt + \sigma_i dz_i \tag{7.6}$$

Expected returns and variances of returns are also assumed to follow an Ito process:

$$d\mu_i = a_i dt + b_i dq_i, \tag{7.7}$$

$$d\sigma_i = f_i dt + g_i dx_i, \tag{7.8}$$

where dq_i and dx_i are two, not necessarily independent, Wiener processes. Equations (7.6)-(7.8) are assumed to form a joint Markov process.

For the further analysis we assume that we have n risky assets and one riskless asset, numbered $n+1$. The riskless rate of return can change over time, i.e. we have

$$\sigma_{n+1} = 0, \tag{7.9}$$

$$\mu_{n+1} = r, \tag{7.10}$$

$$b_{n+1} \neq 0. \tag{7.11}$$

Unlike in the CAPM the investor does not maximize his terminal utility at the end of a given time period (his time horizon), but the utility over the whole time period with length $T > 0$.[6] The investor will maximize the function

$$E\left[\int_0^T U^k\left(C^k(t)\right)e^{-\rho^k t}dt + U^k\left(W^k(T)\right)e^{-\rho^k T}|\Omega_0\right], \tag{7.12}$$

where U^k denotes the utility of individual k, C^k its consumption at time t and ρ the discount factor for future utility. The first term denotes the present value

[6] We assume that utility can only come from consumption, investment and the existence of wealth are not part of the utility. The utility from wealth originates from the possibility of future consumption. A similar distinction has indirectly been made by Adam Smith. See Smith (1776, p.47).

of consumption from time 0 to time$T$ and the second term denotes the present value of the utility from terminal wealth $W^k(T)$, or terminal consumption.

Merton (1990, pp. 124 ff.) derives the dynamics for the wealth $W^k(t)$ at any point of time. It consists of the number of assets that have been hold at a previous point of time, $N_i^k(t-h)$ multiplied by its present price, P_t^i:

$$W^k(t) = \sum_{i=1}^{n+1} N_i^k(t-h) P_t^i. \tag{7.13}$$

In the same moment of time he chooses a new portfolio, i.e. a new number of shares, $N_i^k(t)$ and the optimal consumption per unit of time, $C^k(t)h$. His wealth becomes[7]

$$W^k(t) = \sum_{i=1}^{n+1} N_i^k(t) P_t^i + C^k(t)h. \tag{7.14}$$

Solving (7.13) and (7.14) for $C^k(t)h$ gives

$$C^k(t)h = \sum_{i=1}^{n+1} \left(N_i^k(t-h) - N_i^k(t)\right) P_t^i \tag{7.15}$$

or with shifting the time period by h:

$$C^k(t+h)h = \sum_{i=1}^{n+1} \left(N_i^k(t) - N_i^k(t+h)\right) P_{t+h}^i \tag{7.16}$$

$$= \sum_{i=1}^{n+1} \left(N_i^k(t) - N_i^k(t+h)\right) \left(P_{t+h}^i - P_t^i\right) + \sum_{i=1}^{n+1} \left(N_i^k(t) - N_i^k(t+h)\right) P_t^i \tag{7.17}$$

By letting $h \to 0$ (6.16) and (7.13) become

$$W^k(t) = \sum_{i=1}^{n+1} N_i^k(t) P_t^i, \tag{7.18}$$

$$-C^k(t)dt = \sum_{i=1}^{n+1} dN_i^k(t) dP_t^i + \sum_{i=1}^{n+1} dN_i^k(t) P_t^i. \tag{7.19}$$

[7] By the assumption that the price does not change with his trading decision.

Differentiating (7.18) by using Ito's lemma gives us with (7.19)

$$\begin{aligned} dW^k(t) &= \sum_{i=1}^{n+1} N_i^k(t) dP_t^i + \sum_{i=1}^{n+1} dN_i^k(t) dP_t^i + \sum_{i=1}^{n+1} dN_i^k(t) P_t^i \\ &= \sum_{i=1}^{n+1} N_i^k(t) dP_t^i - C^k(t) dt. \end{aligned} \tag{7.20}$$

The first term represents the capital gains made from investing into assets and the last term are the losses in wealth due to consumption.[8]

Define $w_i^k(t) = \frac{N_i^k(t) P_t^i}{W^k(t)}$ as the fraction of wealth invested into asset i after consumption. Inserting this definition into (7.20) we get

$$dW^k(t) = \sum_{i=1}^{n+1} w_i^k(t) \frac{dP_t^i}{P_t^i} W^k(t) - C^k(t) dt. \tag{7.21}$$

Inserting (7.6) gives us the Ito process governing wealth:

$$\begin{aligned} dW^k(t) &= \sum_{i=1}^{n+1} w_i^k(t) \mu_i W^k(t) dt \\ &\quad + \sum_{i=1}^{n+1} w_i^k(t) \sigma_i W^k(t) dz_i - C^k(t) dt \\ &= W^k(t) \left[\sum_{i=1}^{n+1} w_i^k(t)(\mu_i - r) + r \right] dt \\ &\quad - C^k(t) dt + W^k(t) \sum_{i=1}^{n+1} w_i^k(t) \sigma_i dz_i. \end{aligned} \tag{7.22}$$

Define further $X = (X_1, \ldots, X_m)$ to be the vector of state variables, which we assume to follow an Ito process:

$$dX = F dt + G dQ. \tag{7.23}$$

where $dQ = (dq_q, \ldots, dq_m)$ is a Wiener process, with ν_{ij} denoting the correlation between dq_i and dq_j, η_{ij} between dq_i and dz_j. It is further $F = (f_1, \ldots, f_m)$ and $G = (g_1, \ldots, g_m)$.

[8] We could include other sources of income than capital gains, e.g. wages, that would increase wealth in every period of time. This inclusion would not change the argumentation and is therefore omitted here for simplicity.

After these preliminaries we now can solve the maximization problem of equation (7.12). For determining the maximum we have to find the optimal amount of consumption, $C^k(t)$ in every moment of time and the optimal share of the remaining wealth to invest into each asset, $\{w_i^k(t)\}_{i=1}^{n+1}$.[9] We define a performance function $J^k(W^k, t, X)$ as[10]

$$J^k(W^k, t, X) \equiv \max_{C^k(t), \left\{w_i^k(t)\right\}_{i=1}^n} E\left[\int_t^T U\left(C^k(\tau)\right) e^{-\rho^k \tau} d\tau + U\left(W^k(T)\right) e^{-\rho^k T} | \Omega_t\right]. \tag{7.24}$$

As

$$J^k(W^k, T, X) = U\left(W^k(T)\right) e^{-\rho^k T} \tag{7.25}$$

we can rewrite (7.24) as

$$J^k(W^k, t, X) = \max_{C^k(t), \left(w_i^k(t)\right)_{i=1}^n} E\left[\int_t^T U\left(C^k(\tau)\right) e^{-\rho^k \tau} d\tau + J^k(W^k, T, X) | \Omega_t\right]. \tag{7.26}$$

With $v \equiv T - t$ applying the mean-value theorem states that there exists a $t^* \in [t, T]$ such that

$$\int_t^T U\left(C^k(\tau)\right) e^{-\rho^k \tau} d\tau = U\left(C^k(t^*)\right) e^{-\rho^k t^*} v. \tag{7.27}$$

We can expand $J^k(W^k, T, X)$ in a second order Taylor series around (W^k, t, X):

$$\begin{aligned} J^k(W^k, T, X) \approx\ & J^k(W^k, t, X) + J_W^k \Delta W + J_t^k v + J_X^k \Delta X \\ & + \frac{1}{2} J_{WW}^k \Delta W^2 + J_{WX}^k \Delta W \Delta X + \frac{1}{2} J_{tt}^k v^2 \\ & + J_{Wt}^k \Delta W v + J_{Xt}^k \Delta X v + \frac{1}{2} J_{XX}^k \Delta X^2, \end{aligned} \tag{7.28}$$

where the subindices denote the derivative with respect to this variable evaluated at $(W^k(t), t, X(t))$, $\Delta W = W^k(T) - W^k(t)$ and $\Delta X = X(T) - X(t)$. For

[9] As $\sum_{i=1}^{n+1} w_i^k(t) = 1$ it is sufficient to determine the fractions of the first n risky assets, the fraction to be invested into the $(n+1)$th riskless asset is therewith also determined.

[10] For the method of solving this problem with principles of dynamic programming see Merton (1973, pp. 875 ff.), Merton (1990, pp. 100 ff.) and Intriligator (1971, pp. 326 ff.).

simplicity we assume that the derivatives J^k_W and J^k_X do not vary over time, i.e. $J^k_{Wt} = J^k_{Xt} = 0$, which is resonable if we assume v to be not too large. Inserting (7.27) and (7.28) into (7.26), eliminating $J^k(W^k, t, X)$ and dividing by v gives us:

$$(7.29) \qquad \max_{C^k(t), \left\{w^k_i(t)\right\}^n_{i=1}} \quad E\Bigg[U\left(C^k(t^*)\right)e^{-\rho^k t^*} + J^k_W \frac{\Delta W}{v} + J^k_t + J^k_X \frac{\Delta X}{v} + \frac{1}{2}J^k_{WW}\frac{\Delta W^2}{v} + J^k_{WX}\frac{\Delta W \Delta X}{v} + \frac{1}{2}J^k_{tt}v + \frac{1}{2}J^k_{XX}\frac{\Delta X^2}{v}|\Omega_t\Bigg] = 0.$$

Taking the limit as $v \to 0$, hence $t^* \to t$, we get as W^k and X follow a continuous Ito process:

$$(7.30) \qquad \max_{C^k(t), \left\{w^k_i(t)\right\}^n_{i=1}} \quad \Bigg[U\left(C^k(t)\right)e^{-\rho^k t} + J^k_W E\left[\frac{dW}{dt}\right] + J^k_t + J^k_X E\left[\frac{dX}{dt}\right] + \frac{1}{2}J^k_{WW}E\left[\frac{dW^2}{dt}\right] + J^k_{WX}E\left[\frac{dWdX}{dt}\right] + \frac{1}{2}J^k_{XX}E\left[\frac{dX^2}{dt}\right]\Bigg] = 0.$$

With the continuity of the Ito process we get $E[dW] = E[dX] = 0$ and therewith $E[dW^2] = Var[dW]$ and $E[dX^2] = Var[dX]$. with the notations of (7.22) and (7.23) we find that

$$\begin{aligned}
J^k_X E\left[\frac{dX}{dt}\right] &= \sum_{i=1}^{m} J^k_i f_i, \\
J^k_{XX} E\left[\frac{dX^2}{dt}\right] &= \sum_{i=1}^{m}\sum_{j=1}^{m} J_{ij} g_i g_j \nu_{ij}, \\
J^k_{WX} E\left[\frac{dWdX}{dt}\right] &= \sum_{i=1}^{m}\sum_{j=1}^{n} J_{iW} w^k_j g_i \sigma_j \eta_{ij} W^k \nu_{ij}, \\
J^k_{WW} E\left[\frac{dW^2}{dt}\right] &= J_{WW}\sum_{i=1}^{n}\sum_{j=1}^{n} w^k_i w^k_j \sigma_{ij} W^2,
\end{aligned}$$

where the subindices i and j are for the derivation with respect to the i's and j's state variable. We further have

$$\frac{dW^k}{dt} = W^k\left(\sum_{i=1}^{n+1} w^k_i(\mu_i - r) + r\right) - C^k.$$

Inserting these results into (7.30) gives

$$\max_{C^k(t),\left\{w_i^k(t)\right\}_{i=1}^n} \Bigg[U(C^k)e^{-\rho^k t} + J_t^k + J_W^k\left(W^k\left(\sum_{i=1}^{n+1} w_i^k(\mu_i - r) + r\right) - C^k\right) + \sum_{i=1}^m j_i^k f_i + \frac{1}{2}J_{WW}\sum_{i=1}^m\sum_{j=1}^n w_i^k w_j^k \sigma_{ij} W^2 + \sum_{i=1}^m\sum_{j=1}^n J_{iW} w_j^k g_i \sigma_j \eta_{ij} W^k \nu_{ij} + \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m J_{ij} g_i g_j \nu_{ij}\Bigg] = 0. \tag{7.31}$$

Defining the function in brackets for notational simplicity as Φ and denoting the derivitaive with respect to C^k as Φ_C and with respect to w_i^k as Φ_i we get the $(n+1)$ first order conditions for a maximum:[11]

$$\Phi_C = U_C - J_W^k = 0, \tag{7.32}$$

$$\Phi_i = J_w^k(\mu_i - r)W^k + J_{WW}^k \sum_{j=1}^n w_j^k \sigma_{ij} W^2 + \sum_{j=1}^m J_{jW}^k W^k g_j \sigma_i \eta_{ij} = 0. \tag{7.33}$$

Condition (7.32) states the usual result that marginal utility of immediate consumption (U_C) has to equal marginal utility from deferred consumption (J_W).

If we define $V = \{\sigma_{ij}\}_{i,j=1}^n$ as the covariance matrix of the assets, $d^k = \{w_j W^k\}_{j=1}^n$ as the demand vector for the assets of investor k, $\mu = \{\mu_j\}_{j=1}^n$ as the vector of expected returns, $\sigma = \{g_j \sigma_i \eta_{ij}\}_{i,j=1}^{n,m}$ as the covariance matrix of the asset returns and the state variables, $A^k = -\frac{J_W^k}{J_{WW}}$, $H_j^k = -\frac{J_{jW}^k}{J_{WW}^k}$ and $H^k = \{H_i^k\}_{j=1}^n$ we can rewrite (7.34) after dividing by W^k in vector form:[12]

$$-A^k(\mu - r\iota) + Vd^k - H^k\sigma = 0. \tag{7.34}$$

[11] The second order conditions are not reported here, but they can be shown to fulfill the requirements for a maximum.

[12] See Merton (1990, pp. 370 ff.).

Solving for the demand we get:[13]

$$d^k = A^k V^{-1}(\mu - r\iota) + H^k V^{-1}\sigma \tag{7.35}$$

With v_{ij} as the $(i,\, j)$th element of V^{-1} and $\sigma^*_{ij} = g_j \sigma_i \eta_{ij}$ we can rewrite (7.35) for an individual asset as

$$d^k_i = A^k \sum_{j=1}^{n} v_{ij}(\mu_j - r) + \sum_{j=1}^{m}\sum_{l=1}^{n} H^k_l \sigma^*_{jl} v_{ij}. \tag{7.36}$$

The demand for an asset consists of two parts. The first term is the demand similar to that of the static CAPM, it is due to an efficient investment by mean-variance maximizing. A^k includes a term for risk aversion of the investor and $\mu_j - r$ is the risk premium.Tthe second term is an adjustment made for hedging against an unfavorable shift in the state variables. This adjustment can be either positive or negative, depending on the sign of $\frac{\partial C^k}{\partial x_i}$, which forms a part of H^k_i.[14]

If the state variables do not vary over time, it turns out that $H^k_l = 0$ and the results are identical to the standard CAPM.[15] In general, however, the state variables will change over time. For simplicity we now assume that only one state variable varies over time, e.g. the risk free interest rate r. In this case (7.36) becomes

$$d^k_i = A^k \sum_{j=1}^{n} v_{ij}(\mu_j - r) + H^k_r \sum_{j=1}^{m} \sigma^*_{jr} v_{jr}. \tag{7.37}$$

Assume there exists an asset, the nth asset, that is perfectly negative correlated with the state variable, i.e. $\rho_{nr} = -1$.[16] Such an asset could be a bond, which is riskless in terms of default and whose value only depends on the interest rate. Because of the varying interest rate the bond would no longer be riskless in his

[13] As long as the risky assets are not linear dependent, e.g. portfolios are included, the covariance matrices will be non-singular.

[14] See Merton (1973, p. 876). The results of the conditional CAPM that also adjusts for the risk of a change in the beta could be interpreted in the same way.

[15] See Merton (1973, pp. 877 ff.).

[16] If such an asset does not exist, we can consider the asset with the highest correlation in absolute value without changing the results, see Breeden (1979, p. 272). The negative correlation is only taken for computational simplicity, by holding the asset short a perfect positive correlation can be obtained.

value. We find

$$\rho_{jr} = \rho_{jn}\rho_{nr} = -\rho_{jn}, \tag{7.38}$$

such that

$$\sigma^*_{jr} = \rho_{jr}\sigma_j g_r = -\sigma_{jn}\sigma_j g_r = -g_r \frac{\rho_{jn}\sigma_j\sigma_n}{\sigma_n} = -g_r \frac{\sigma_{jn}}{\sigma_n}. \tag{7.39}$$

By using (7.39) the second term in (7.37) becomes $-H^k_r \frac{g_r}{\sigma_n} \sum_{j=1}^n \nu_{ij}\sigma_{jn}$. Because ν_{rj} is an element of the inverse of σ we find that if $i = n$

$$\sum_{j=1}^{n} v_{ij}\sigma_{jn} = 1 \tag{7.40}$$

and otherwise

$$\sum_{j=1}^{n} v_{ij}\sigma_{jn} = 0. \tag{7.41}$$

Therewith we can simplify (7.37) to

$$\begin{aligned} d^k_i &= A^k \sum_{j=1}^{n} v_{ij}(\mu_j - r) \quad \text{for } i = 1, \ldots, n-1, \\ d^k_n &= A^k \sum_{j=1}^{n} v_{nj}(\mu_j - r) - H^k_r \frac{g}{\sigma_n}. \end{aligned} \tag{7.42}$$

We define now three portfolios:[17]

- **Portfolio 1**: Consists of all n risky assets with weights $\delta_i = \frac{\sum_{j=1}^n v_{ij}(\mu_j - r)}{\sum_{j=1}^n \sum l=1^n v_{jl}(\mu_j - r)}$, which are equal to the weights of the optimal risky portfolio in portfolio theory.
- **Portfolio 2**: Consists only of the nth asset.
- **Portfolio 3**: Consists only of the riskless asset.

By construction the compositions of the portfolios do not depend on the preferences of the investors. Let λ^k_i $(i = 1, 2, 3)$ denote the fraction of his total wealth

[17] See MERTON (1973, pp. 880 f.).

investor k invests into each of these portfolios. The total demand for the first $n-1$ assets is given by $\lambda_1^k \delta_i W^k$ and for the nth asset by $(\lambda_1^k \delta_n + \lambda_2^k)W^k$. These demands have to satisfy the conditions (7.42) if an investor has to be indifferent between investing into all n assets directly or investing only in the three portfolios:

$$\lambda_1^k \delta_i W^k = \lambda_1^k \frac{\sum_{j=1}^n v_{ij}(\mu_j - r)}{\sum_{j=1}^n \sum l = 1^n v_{jl}(\mu_j - r)} W^k = d_i^k = A^k \sum_{j=1}^n v_{ij}(\mu_j - r),$$

$$(\lambda_1^k \delta_n + \lambda_2^k)W^k = d_n^k = A^k \sum_{j=1}^n v_{nj}(\mu_j - r) - H_r^k \frac{g}{\sigma_n}.$$

Solving for λ_1^k and λ_2^k gives

$$\lambda_1^k = \frac{A^k}{W^k} \sum_{j=1}^n \sum_{l=1}^n v_{ij}(\mu_j - r), \tag{7.43}$$

$$\lambda_2^k = -\frac{H_r^k}{W^k} \frac{g_r}{\sigma_n}. \tag{7.44}$$

Instead of directly investing into all n risky assets, an investor will be indifferent to choosing a combination of the optimal risky portfolio and the nth asset that has the highest correlation with the changing state variable. The remaining wealth is invested into the third portfolio consisting only of the riskless asset. This result is known as the *Three-Fund Theorem*. It is the dynamic equivalent of Tobin's Separation Theorem.

Portfolios one and three ensure the investor to have a mean-variance efficient holding of assets, i.e. his holdings are on the efficient frontier of the CAPM. But as the state variable changes over time stochastically, the efficient frontier shifts over time, portfolio two hedges against an unfavorable shift.

In the general case where $m \geq 1$ state variables can change over time, it has been shown in MERTON (1990, pp. 499 ff.) that $(m+2)$ portfolios are combined by the investors, hence it is known as the *(m+2)-Fund Theorem*. The properties in the general case do not change, one portfolio remains to be the optimal risky portfolio, one consists only of the riskless assets and the other m portfolios consist only of the asset having the highest correlation with one of the state variables.

We can now derive the equilibrium expected returns in the case of $m = 1$ state variables. Solving (7.34) for $\mu - r\iota$ gives

$$\mu - r\iota = \frac{1}{A^k} V d^k - \frac{H_r^k}{A^k}\sigma. \tag{7.45}$$

As this relation holds for every investor individually, it also has to remain valid when aggregating the demands of all K investors. Let $A = \sum_{K=1}^{K} A^k$, $H = \sum_{K=1}^{K} H^k$, $D_i = \sum_{K=1}^{K} d_i^k$ and $D = \{D_i\}_{i=1}^{n}$. We can rewrite (7.45) as

$$\mu - r\iota = \frac{1}{A} V D - \frac{H}{A}\sigma. \tag{7.46}$$

As we have assumed that trade only occurs in equilibrium, the total demand always has to equal the value of all assets, M: $\sum_{i=1}^{n+1} D_i = M$. We define $w_i = \frac{D_i}{M}$ as the fraction of the ith asset in the total market value, which because of the equilibrium assumption has to equal relative demand. In scalar form (7.46) becomes for $i = 1, \ldots, n$:

$$\begin{aligned} \mu_i - r &= \frac{M}{A}\sum_{j=1}^{n} w_j \sigma_{ij} - \frac{H}{A}\sigma_{ir} \\ &= \frac{M}{A}\sum_{j=1}^{n} w_j \sigma_{ij} + \frac{H}{A}\frac{\sigma_{in} g_r}{\sigma_n}. \end{aligned} \tag{7.47}$$

If we denote the market portfolio by the subindex M we have $\mu_M - r = \sum_{j=1}^{n} w_j(\mu_j - r)$, $\sigma_{iM} = \sum_{j=1}^{n} w_j \sigma_{ij}$ and $\sigma_M^2 = \sum_{j=1}^{n} w_j \sigma_{jM}$ and (7.47) becomes

$$\mu_i - r = \frac{M}{A}\sigma_{iM} - \frac{H}{A}\sigma_{ir}. \tag{7.48}$$

Multiplying by w_j and summing over all $i = 1, \ldots, n$ gives

$$\mu_M - r = \frac{M}{A}\sigma_M^2 + \frac{H}{A}\frac{\sigma_{Mn} g_r}{\sigma_n} = \frac{M}{A}\sigma_M^2 - \frac{H}{A}\sigma_{Mr}. \tag{7.49}$$

For $i = n$ (7.47) becomes

$$\mu_n - r = \frac{M}{A}\sigma_{nM} - \frac{H}{A}\sigma_{nr}. \tag{7.50}$$

By solving (7.49) and (7.50) for $\frac{M}{A}$ and $\frac{H}{A}$ we get

$$\frac{M}{A} = \frac{\sigma_{nr}}{\sigma_{nr}\sigma_M^2 - \sigma_{Mr}\sigma_{nM}}(\mu_M - r) - \frac{\sigma_{Mr}}{\sigma_{nr}\sigma_M^2 - \sigma_{Mr}\sigma_{nM}}(\mu_n - r), \tag{7.51}$$

$$\frac{H}{A} = \frac{\sigma_{nM}}{\sigma_{nr}\sigma_M^2 - \sigma_{Mr}\sigma_{nM}}(\mu_M - r) - \frac{\sigma_M^2}{\sigma_{nr}\sigma_M^2 - \sigma_{Mr}\sigma_{nM}}(\mu_n - r). \tag{7.52}$$

Inserting these results into (7.47) gives us the equilibrium expected returns after rearranging terms:

$$\begin{aligned}
\mu_i - r &= \frac{\sigma_{nr}\sigma_{iM} - \sigma_{nM}\sigma_{ir}}{\sigma_{nr}\sigma_M^2 - \sigma_{Mr}\sigma_{nM}}(\mu_M - r) \\
&\quad + \frac{\sigma_M^2\sigma_{ir} - \sigma_{Mr}\sigma_{iM}}{\sigma_{nr}\sigma_M^2 - \sigma_{Mr}\sigma_{nM}}(\mu_n - r) \\
&= \beta_i^1(\mu_M - r) + \beta_i^2(\mu_n - r),
\end{aligned} \tag{7.53}$$

with $\beta_i^1 = \frac{\sigma_{nr}\sigma_{iM} - \sigma_{nM}\sigma_{ir}}{\sigma_{nr}\sigma_M^2 - \sigma_{Mr}\sigma_{nM}}$ and $\beta_i^2 = \frac{\sigma_M^2\sigma_{ir} - \sigma_{Mr}\sigma_{iM}}{\sigma_{nr}\sigma_M^2 - \sigma_{Mr}\sigma_{nM}}$.

The excess returns compensate investors for the systematic risk (β_i^1) like in the CAPM and additionally for the risk of an unfavorable shift in the state variable (β_i^2). Even assets which have a beta of zero, i.e. are uncorrelated with the market portfolio, may have returns higher than the riskless rate of return because of the exposure to an unfavorable shift in the state variable. This relation can be observed in reality and could not be explained by the CAPM.

By allowing in general $m \geq 1$ state variable to change over time, Merton (1990, p. 510) has shown that (7.53) in general becomes

$$\mu_i - r = \beta_i^1(\mu_M - r) + \sum_{j=1}^{m} \beta_i^j(\mu_j - r). \tag{7.54}$$

i.e. every risk of a state variable is compensated individually. The ICAPM extends the CAPM to a dynamic environment. The results are similar to those of the APT, but it has the advantage that the risks can be determined from characteristics of the assets. Cox et al. (1985) derive a similar result in a much more general framework than provided here.

7.2 Empirical evidence

Empirical investigations into the performance of the ICAPM are not frequently found. The reason is that at a first glance the ICAPM and the APT look alike from their results. The ICAPM only has a fixed risk factor, the market portfolio. In the APT the market portfolio is not a risk factors as it has been assumed that an investor is perfectly well diversified and hence the only source of risk are common factors. In the ICAPM portfolios have not to be perfectly diversified, nor the market portfolio neither the portfolios having the highest correlation with the state variables, hence we could interpret the risk from the market portfolio as arising from imperfect diversification. If all portfolios are perfectly diversified and the state variables equal the common factors, the ICAPM collapses to the APT. We could therefore view the APT as a special case of the ICAPM.[18] Therefore the APT and ICAPM are often treated alike, despite their different theoretical foundations.[19]

The investigation by DOKKO/EDELSTEIN (1991) explicitly uses the ICAPM as their reference model. They use the uncertainty about future inflation and real production as state variables. By using monthly data for the period 1960-1985 they show that excess returns are significantly affected by changes in these state variables.

They further find changes inflation uncertainty to be highly persistent, whereas changes in real production uncertainty are only temporary. Therefore they find that changes in inflation uncertainty influence asset prices significantly, whereas real production uncertainty has no important influence.[20]

With these results DOKKO/EDELSTEIN (1991) can explain the behavior of the stock market in their observed time period quite well. But as most models it fails to explain short-term movements of prices.

The ICAPM faces the same problem of identifying state variables as does one in the APT with identifying common factors. FAMA (1998) addresses this problem

[18] See FAMA (1996, pp. 460 f.).
[19] E.g. CONSTANTINIDES/MALLIARIS (1995) do so.
[20] This result is due to the finding of the present value model where has been pointed out that a permanent change affects the price much more than a temporary change.

of determining state variables.

8. The Consumption-Based Capital Asset Pricing Model

When we want to apply the ICAPM to explain the behavior of asset prices we face the problem of identifying the relevant state variables, the theory does give no hint how to choose the relevant variables. The *Consumption-Based Capital Asset Pricing Model* (*CCAPM*) as first developed by Breeden (1979) develops the ICAPM further within the same theoretical framework to aggregate the risks from shifting state variables into a single variable, consumption.

8.1 Derivation of the model[1]

The assumptions underlying the CCAPM are identical to those of the ICAPM developed in the last section. The investors are also assumed to maximize the function already stated in equation (7.12)[2]

$$E\left[\int_0^T U^k\left(C^k(t)\right)e^{-\rho^k t}dt + U^k\left(W^k(T)\right)e^{-\rho^k T}|\Omega_0\right] \tag{8.1}$$

with respect to consumption $C^k(t)$ and the portfolio composition $\{w_i^k\}_{i=1}^n$.

The only additional assumption we have to make concerns consumption. We assume that there exists only only a single consumption good[3] and that the state variables influence also the consumption, i.e. consumption becomes a stochastic variable. This influence could e.g. be through influencing the price and therewith

[1] The presentation follows Breeden (1979).
[2] The notation is identical to that of the last section on the ICAPM.
[3] The results remain valid with many consumption goods. The only changes that have to be made are to replace nominal variables by real variables. See Breeden (1979, pp. 285 ff.).

the number of goods that can be bought with a given wealth.

The first order conditions for maximizing (8.1) have already been derived in section 4.7. We restate them here for convenience slightly modified:

$$J_W^k = U_C, \tag{8.2}$$

$$d^k = w^k W^k = -\frac{J_W^k}{J_{WW}^k} V^{-1}(\mu - r\iota) - \frac{J_{XW}^k}{J_{WW}^k} V^{-1}\sigma. \tag{8.3}$$

If we derive (8.2) again with respect to the state variables X and wealth W^k we get

$$J_{WX}^k = U_{CX} = U_{CC} C_X^k, \tag{8.4}$$

$$J_{WW}^k = U_{CW} = U_{CC} C_W^k. \tag{8.5}$$

Inserting these results into (8.3) we get

$$\begin{aligned} w^k W^k &= -\frac{U_C}{U_{CC} C_W^k} V^{-1}(\mu - r\iota) - \frac{U_{CC} C_X^k}{U_{CC} C_W^k} V^{-1}\sigma \\ &= z^k \frac{1}{C_W^k} V^{-1}(\mu - r\iota) - \frac{C_X^k}{C_W^k} V^{-1}\sigma, \end{aligned} \tag{8.6}$$

with $z^k = -\frac{U_C}{U_{CC}}$ as the Arrow-Pratt measure of risk aversion. Premultiplying (8.6) by $C_W^k V$ gives

$$C_W^k V w^k W^k = z^k(\mu - r\iota) - C_X^k \sigma. \tag{8.7}$$

The term $V w^k W^k$ represents the vector of covariances of the asset returns with a change in the wealth. We will denote $V_{\mu W^k} \equiv V w^k W^k$. Rearranging (8.7) gives

$$z^k(\mu - r\iota) = V_{\mu W^k} C_W^k + C_X^k \sigma. \tag{8.8}$$

The covariances of the asset returns with the investors consumption can be derived using Ito's lemma:

$$V_{\mu C^k} = V_{\mu W^k} C_W^k + C_X^k \sigma. \tag{8.9}$$

Inserting (8.9) into (8.8) gives

$$z^k(\mu - r\iota) = V_{\mu C^k}. \tag{8.10}$$

As (8.10) holds for every investor we can aggregate their decisions. Defining $z \equiv \sum_{k=1}^{K} z^k$ and $V_{\mu C} \equiv \sum_{k=1}^{K} V_{\mu C^k}$[4] gives us

$$z(\mu - r\iota) = V_{\mu C}, \tag{8.11}$$

or for a single asset

$$z(\mu_i - r) = \sigma_{iC}. \tag{8.12}$$

Equation (8.12) has also to be fulfilled for any portfolio. Let such a portfolio be denoted by the subindex M, it has not necessarily to be the market portfolio. With the weights w_i^M of this portfolio we can define the portfolio's expected return and covariance with consumption as $\mu_M \equiv \sum_{i=1}^{n} w_i^M \mu_i$ and $\sigma_{MC} = \sum_{i=1}^{n} w_i^M \sigma_{iC}$. Multiplying (8.12) with w_i^M and adding over all assets gives us

$$z(\mu_M - r) = \sigma_{MC}. \tag{8.13}$$

Solving this equation for z and inserting into (8.12) gives after rearranging

$$\mu_i - r = \frac{\sigma_{iC}}{\sigma_{MC}}(\mu_M - r). \tag{8.14}$$

By defining betas with respect to aggregate consumption in the conventional way as $\beta_{iC} \equiv \frac{\sigma_{iC}}{\sigma_C^2}$ and $\beta_{MC} \equiv \frac{\sigma_{MC}}{\sigma_C^2}$ we can rewrite (8.14) as

$$\mu_i - r = \frac{\beta_{iC}}{\beta_{MC}}(\mu_M - r). \tag{8.15}$$

Suppose there exists a portfolio M that is perfectly correlated with changes in aggregate consumption, then (8.15) reduces to[5]

$$\mu_i - r = \beta_{iC}(\mu_M - r). \tag{8.16}$$

The risks of an unfavorable shift in the state variables that determined the excess returns in the ICAPM have now been aggregated into a general risk factor,

[4] We assume here reasonably that the investors make their decisions independent of each other, hence we can neglect covariances in decisions.

[5] The same result can be obtained when choosing the portfolio with the highest correlation, see Breeden (1979, p. 276). It is not necessary to determine a market portfolio and face the therewith associated problems as stated in the CAPM, but it is possible to determine any portfolio as a reference.

aggregate consumption. The different risk factors all influence current and future consumption, hence they can be aggregated in this way. The term $\mu_M - r$ is also called the *market price of consumption risk*.

In order to interpret this result we have to remember that the present utility, $J^k(W^k, t, X)$ is, beside terminal wealth, affected by the amount of consumption and its timing. As we further can reasonably argue that $C^k_W > 0$, i.e. the higher the wealth the higher consumption, and wealth is affected by these returns, consumption is also affected by the returns of the assets held.

Assume now that a state occurs that reduces current consumption and the expected return of an asset, i.e. $\beta_{iC} \gg 0$. The reduced current consumption reduces the investors present utility. Because expected future returns of the asset are also reduced, future expected wealth is reduced and hence future expected consumption, what reduces utility further. When holding the asset an investor faces this risk of an unfavorable shift in the state variables that influence his future consumption, hence he has to be compensated for this risk. The same argumentation can be used for a negative beta.

8.2 Empirical investigations

An empirical investigation of the CCAPM faces several measurement problems concerning aggregate consumption:[6]

- The statistical data on aggregate consumption capture not consumption directly, but the expenditures for consumption goods and services. As the goods purchased are not necessarily consumed immediately (e.g. they can be stored for later consumption or can be consumed over time like consumer durables), the data will be biased, although as a result of aggregation the bias will be reduced, but it may lead or lag the business cycle.

[6] See Breeden et al. (1989, pp. 234 ff.).

- The data available are not for an instant of time, but denote consumption for a certain period, at least a month or a quarter, whereas asset prices are available on a daily or even intraday basis.

- The data are generated from samples and we face the problem of sampling errors.

Compared to the use of a market portfolio as in the CAPM the use of consumption data also has a virtue. While the market portfolio that is determined typically does not include important assets like real estate or human capital, consumption data cover a much larger fraction of the effective consumption.[7]

Beside the data problem an empirical investigation faces many econometric problems, e.g. the determination of the portfolio with the highest correlation with a change in aggregate consumption (also called the *Maximum Correlation Portfolio*), which is the equivalent to the market portfolio in the CAPM. For an overview of these econometric considerations see Breeden et al. (1989).

Breeden et al. (1989) find that excess returns of a zero beta portfolio are small, as predicted by the theory, whereas they are relatively large in the CAPM model. They report further that the market price of consumption risk is positive by observing data for the period 1929-1982 using quarterly and monthly data. The linearity of excess returns and consumption risk (β_{iC}) is only rejected for more recent subperiods (1947-1982) where the data quality is improved. This suggests that data quality is a crucial factor for interpreting the results.

Summarizing can be said that the results show a weak support for the CCAPM, a situation similar to the results concerning the CAPM. As other models presented thus far the CCAPM is especially not able to explain the behavior of asset prices in the short run, but this may also be the consequence of missing consumption data for periods below a month.

[7] See Breeden (1979, pp. 291 ff.).

AHN/CHO (1991) show that with a time varying risk aversion, where the risk aversion depends on past returns, the CCAPM becomes much more consistent with the data.

9. The International Capital Asset Pricing Model

The previous models of asset pricing implicitly assumed that all investors are located in the same country and consider only assets in their home country. In reality, however, we find investors in different countries and they also invest a part of their wealth abroad. In such a framework we therefore should consider the influence of exchange rates, different tastes for consumption across countries and barriers to foreign investment. These points were incorporated into the *International Capital Asset Pricing Model* (*International CAPM*) as developed by Stulz (1981b), Stulz (1981a) and Stulz (1995).

9.1 No differences in consumption and no barriers to foreign investment

Stulz (1995) provides a model where neither the investors have different tastes across countries nor different costs of investing at home and abroad are faced. The assumptions underlying this model are very similar to those of the ICAPM, they are listed in table 9.1.

We assume that the price of asset i, denominated in the currency of country j, P_{ij} follows an Ito process:

$$\frac{dP_{ij}}{P_{ij}} = \mu_{ij}dt + \sigma_{ij}dz_{ij}, \tag{9.1}$$

where μ_{ij} denotes the expected nominal return of asset i in currency j, σ_{ij} the standard deviation of this return and dz_{ij} a standard Wiener process. The price

- Asset prices and exchange rates are jointly log-normal distributed
- There exists a single consumption good or equivalently a common basket of consumption goods
- The consumption good is continuously and costless traded between the countries
- There are n risky assets that are traded continuously and costless
- There is an asset that is riskless in real terms, i.e. in terms of the consumption good
- There are J countries
- There exist no transaction costs, taxes, transportation costs or tariffs
- No restrictions on short sales
- No barriers to international investment
- Investors are price takers and have the same information
- Investors are risk averse
- For the consumption good the law of one price applies, i.e. the price is equal in all countries adjusted only by the exchange rate (Purchasing Power Parity, the exchange rate changes according to the difference of inflation in the countries.

Table 9.1: Assumptions of the International CAPM

of the consumption good in country j, P_j^C, also follows an Ito process:

$$\frac{dP_j^C}{P_j^C} = \pi_j dt + \sigma_{\pi_j} dz_{\pi_j}, \tag{9.2}$$

where π_j denotes the expected inflation rate in country j, σ_{π_j} the inflation rate's variance and dz_{π_j} a standard Wiener process. The real price of an asset in terms of the consumption good is given by

$$P_{ij}^r = \frac{P_i j}{P_j^C} \tag{9.3}$$

and can be shown by Ito's lemma to follow

$$\frac{dP_{ij}^r}{P_{ij}^r} = \frac{d\frac{P_{ij}}{P_j^C}}{\frac{P_{ij}}{P_j^C}} = \left[\mu_{ij} - \pi_j - \sigma_{ij,\pi_j} + \sigma_{\pi_j}^2\right] dt + \sigma_{ij} dz_{ij} - \sigma_{\pi_j} dz_{\pi_j}, \tag{9.4}$$

where σ_{ij,π_j} denotes the covariance between the nominal asset return and the inflation in this country. The first term denotes the expected real return of asset i in country j, $\mu_{ij}^r \equiv \mu_{ij} - \pi_j - \sigma_{ij,\pi_j} + \sigma_{\pi_j}^2$. As the sum of two Wiener processes is again a Wiener process we define a Wiener process $\sigma_{ij}^r dz_{ij}^r \equiv \sigma_{ij} dz_{ij} - \sigma_{\pi_j} dz_{\pi_j}$ and get an Ito process for the real price of the asset:

$$\frac{dP_{ij}^r}{P_{ij}^r} = \mu_{ij}^r dt + \sigma_{ij}^r dz_{ij}^r. \tag{9.5}$$

We have now formulated exactly the CAPM derived in section 4.3, only substituting nominal returns by real returns. With the law of one price we know that real prices and returns are equal in all countries and we can write (9.5) as

$$\frac{dP_i^r}{P_i^r} = \mu_i^r dt + \sigma_i^r dz_i^r. \tag{9.6}$$

From the CAPM we know that in this case

$$\mu_i^r - r^r = \beta_i^r(\mu_W^r - r^r), \tag{9.7}$$

where the superscript r denotes real variables and μ_W^r is the expected real return on the world market portfolio. In real terms there are no differences between countries and for the decisions it is of no importance where an investor and an asset is located.

It is more convenient to write the returns in nominal than in real returns. Assume therefore that there exists an asset that is riskless in nominal terms[1] and that has a zero beta with the real world portfolio return. This asset also has to fulfill (9.7). By replacing μ_i^r by its original expression we get

$$r_j - \pi_j - \sigma_{ij,\pi_j} + \sigma_{\pi_j}^2 - r^r = 0. \tag{9.8}$$

Rearranging and inserting into (9.7) gives us

$$\mu_{ij} - (\sigma_{ij,\pi_j} - \sigma_{r_j j,\pi_j}) - r_j = \beta_i^r(\mu_W^r - r^r). \tag{9.9}$$

The last expression on the on the right side denotes the real excess return of the world market portfolio. As we know from the derivation of the real excess return this is given by

$$\begin{aligned} \mu_W^r - r^r &= \mu_{Wj} - \pi_j - \sigma_{Wj,\pi_j} + \sigma_{\pi_j}^2 - (r_j - \pi_j - \sigma_{rj,\pi_j} + \sigma_{\pi_j}^2) \\ &= \mu_{Wj} - (\sigma_{Wj,\pi_j} - \sigma_{rj,\pi_j}) - r_j. \end{aligned} \tag{9.10}$$

Stulz (1995, p. 205) makes now the assumption that for all assets $\sigma_{ij,\pi_j} = 0$, i.e. the nominal asset returns are uncorrelated with the inflation rate.[2] With this assumption we see from (9.9) and (9.10) that the real and nominal access returns coincide. The beta of the real returns also equals the beta of the nominal returns, Hence (9.9) becomes

$$\mu_{ij} - r_j = \beta_{ij}(\mu_{Wj} - r_j). \tag{9.11}$$

We therewith found that with the assumption of no differences in tastes and no investment barriers the CAPM as derived in section 4.3 remains valid in an international setting. The market portfolio becomes the world market portfolio. We can view the world as an integrated market, different currencies and borders have no influence with the assumption of Purchasing Power Parity.

[1] This asset is not riskfree in real terms as inflation may reduce its value.

[2] With this assumption the inflation risk has not to be considered and is not compensated. Such an assumption may be reasonable in the present circumstances of the economies in industrialized countries with their low inflation rates, but for countries with high inflation rates this assumption is not reasonable as the inflation affects the economy as a whole. The more reasonable assumption is that inflation affects all assets in the same manner, i.e. $\sigma_{ij,\pi_j} = \sigma_{j,\pi_j}$ for all $i = 1, \ldots, n$. This would give us the same result.

We will further on consider the influence different tastes and investment barriers have on the expected returns of assets.

9.2 Differences in consumption

Stulz (1981a) extends the CCAPM to the case that investors and assets are located in different countries and that the tastes of the investors differ between these countries but are identical within a country. For simplicity we assume that there exist only two countries, the home country and the foreign country, denoted by an asterik at the variables.[3] There exist no barriers to international investment.

To model the different tastes in the countries we assume that there exist k different consumption goods in the home country and k^* in the foreign country and at least one consumption good is consumed in both countries. For the goods consumed in both countries the law of one price applies as before, i.e. with an exchange rate of e between the two currencies we have for these consumption goods:

$$P_j^C = eP_j^{C*}. \tag{9.12}$$

The price for the jth domestic good in domestic currency is assumed to follow an Ito process:

$$\frac{dP_j^C}{P_j^C} = \pi_j dt + \sigma_{\pi_j} dz_{\pi_j}, \tag{9.13}$$

where μ_j^C denotes the expected price change, σ_j^C the variance of the price change and dz_j^C a standard Wiener process. The exchange rate also follows an Ito process:

$$\frac{de}{e} = \mu_e dt + \sigma_e dz_e, \tag{9.14}$$

with μ_e as the expected change in the exchange rate, σ_e its variance and dz_e a standard Wiener process. We assume that there are n risky assets that are traded

[3] The extension of the model to more countries is straightforward and does not change the results.

among countries, hence the law of one price applies to all assets. Additionally each country has a riskless nominal asset in his currency with an interest rate of r and r^*, respectively. These riskless assets are also traded between the countries as the $(n+1)$th asset, but as the exchange rate may vary it is not riskless in the other country.

The risky foreign assets are assumed to follow an Ito process in foreign currency:

$$\frac{dP^*_{i^*}}{P^*_{i^*}} = \mu^*_{i^*}dt + \sigma^*_{i^*}dz_{i^*}, \tag{9.15}$$

where $\mu^*_{i^*}$ is the expected return in foreign currency, $\sigma^*_{i^*}$ the standard deviation of these returns and $dz^*_{i^*}$ a standard wiener process. Using the law of one price for these assets the price in domestic currency is given by $P_{i^*} = eP^*_{i^*}$ and by applying Ito's lemma the dynamics of the foreign assets in domestic currency are given by

$$\frac{dP_{i^*}}{P_{i^*}} = \frac{dP^*_{i^*}}{P^*_{i^*}} + \frac{de}{e} + \frac{de}{e}\frac{dP^*_{i^*}}{P^*_{i^*}}. \tag{9.16}$$

Analogous equations can be derived for the domestic assets. For the riskless assets in domestic currency these dynamics are given by

$$\frac{dP_{n+1}}{P_{n+1}} = rdt, \tag{9.17}$$

$$\frac{dP_{n+1^*}}{P_{n+1^*}} = r^*dt + \frac{de}{e}. \tag{9.18}$$

Define the excess returns of a foreign asset in domestic currency, H_{i^*}, as the return of this asset financed by borrowing abroad at the rate of r^*. We get by using (9.16) and (9.18):

$$\begin{aligned} \frac{dH_{i^*}}{H_{i^*}} &= \frac{dP_{i^*}}{P_{i^*}} - \frac{dP_{n+1^*}}{P_{n+1^*}} \\ &= \frac{dP^*_{i^*}}{P^*_{i^*}} + \frac{de}{e}\frac{dP^*_{i^*}}{P^*_{i^*}} - r^*dt. \end{aligned} \tag{9.19}$$

The dynamics of the exchange rate does not affect the return as the investor has fully hedged this risk by financing his investments abroad. Hence the exchange rate risk will not be compensated as it can be diversified, the only risk is the development of the return in foreign currency. Equation (9.19) for a foreign investor investing into a domestic asset becomes:

$$\frac{dH_i^*}{H_i^*} = \frac{dP_i}{P_i} - \frac{de}{e}\frac{dP_i}{P_i} - rdt. \tag{9.20}$$

For a domestic investor investing at home we have

$$\frac{dH_i}{H_i} = \frac{dP_i}{P_i} - rdt. \tag{9.21}$$

The excess returns of foreign and domestic investors differ only in the term $\frac{de}{e}\frac{dP_i}{P_i}$, i.e. the covariance of the change in the exchange rate and the return of the asset in domestic currency. We further get the excess returns for a domestic investor for the foreign riskless asset by (9.17) and (9.18) as:

$$\begin{aligned} \frac{dH_{n+1^*}}{H_{n+1^*}} &= r^*dt + \frac{de}{e} - rdt \\ &= (r^* - r)dt + \frac{de}{e} \\ \frac{dH_{n+1}^*}{H_{n+1}^*} &= (r^* - r)dt - \frac{de}{e} + \left(\frac{de}{e}\right)^2. \end{aligned} \tag{9.22}$$

After these preliminaries we are now able to derive the model of international asset pricing. Unlike in the CCAPM presented above we now have more than one consumption good. Let q_j^k denote the quantities consumed by domestic investor k of consumption good j, then we have

$$C^k = \sum_{j=1}^{n} P_j^C q_j^k = P^C q^k, \tag{9.23}$$

where C^k denotes the consumption of the kth domestic investor, P^C the vector of domestic consumption good prices and q^k the vector of consumption for each consumption good. Besides the determination of the optimal amount spent for consumption, the investor now additionally has to find the optimal allocation of the different consumption goods.[4] We therefore define the utility function as

$$U^k\left(C^k(P^C),P^C\right) = \max_{q^k} u^k(q^k) = \max_{q^k}\left(u^k(q^k)+\lambda(C^k-P^Cq^k\right). \tag{9.24}$$

The first order conditions for this maximization are

$$u_q = \lambda P^C, \tag{9.25}$$

$$U_P = -\lambda q^k. \tag{9.26}$$

where $\lambda = U_C$ is the shadow price of increased consumption. Hence (9.25) and (9.26) become

$$u_q = U_C P^C, \tag{9.27}$$

$$U_P = -U_C q^k. \tag{9.28}$$

The optimal portfolio for a domestic investor has been derived in equations (8.2) and (8.3) to fulfill the following conditions:

$$J_W^k = U_C, \tag{9.29}$$

$$d^k = w^k W^k = -\frac{J_W^k}{J_{WW}^k}V^{-1}(\mu - r\iota) - \frac{J_{XW}^k}{J_{WW}^k}V^{-1}\sigma. \tag{9.30}$$

We now assume that the first K state variables are the log-prices of consumption goods. Using (9.24) as the utility function and applying the theorem of implicit functions we get from (9.29) by differentiating

$$J_{W,lnP}^k = U_{CC}C_{lnP} + U_{C,lnP}. \tag{9.31}$$

Let Y denote the vector of the state variables that are not prices of consumption goods, i.e. $X = (lnP, Y)'$, therewith we can write J_{XW}^k as

$$J_{XW}^k = \begin{bmatrix} J_{lnPW}^k \\ J_{YW}^k \end{bmatrix} = \begin{bmatrix} U_{CC}C_{lnP} + U_{ClnP} \\ U_{CC}C_Y \end{bmatrix} = U_{CC}C_X + \begin{bmatrix} U_{ClnP} \\ 0 \end{bmatrix} \tag{9.32}$$

[4] These steps up to equation (9.37) follow Breeden (1979, pp. 285 ff.).

by using the results and notations of the CCAPM. Inserting (9.32) into (9.30) and continuing as in 4.8 gives us

$$(9.33)\qquad \begin{aligned} w^k W^k &= -\frac{J^k_W}{J^k_{WW}} V^{-1}(\mu - r\iota) - \frac{U_{CC}C_X}{J^k_{WW}} V^{-1}\sigma \\ &\quad -\frac{1}{J^k_{WW}} \begin{bmatrix} U_{ClnP} \\ 0 \end{bmatrix} V^{-1}\sigma \\ &= \frac{z^k}{C_W} V^{-1}(\mu - r\iota) - \frac{C_X}{C_W} V^{-1}\sigma - \begin{bmatrix} \frac{U_{ClnP}}{U_{CC}C_W} \\ 0 \end{bmatrix} V^{-1}\sigma. \end{aligned}$$

We define α^k_j as the share of total consumption investor k spends for good j

$$(9.34)\qquad \alpha^k_j = \frac{P^C_j q^k_j}{C^k}$$

and m^k_j as the marginal share from an increase in total consumption:

$$(9.35)\qquad m^k_j = P^C_j \frac{\partial q^k_j}{\partial C^k} = P^k_j q^j_C.$$

By using these definitions we get with (9.28):

$$(9.36)\qquad \begin{aligned} U_{C,lnP} &= U_{CP}\frac{\partial P^C}{\partial lnP^C} = U_{PC}P^C = -P^C\left(U_{CC}q^k + U_C q_C\right) \\ &= -U_{CC}P^C q^k - U_C P^C q_C \\ &= -U_{CC}\alpha^k C^k - U_C m^k, \end{aligned}$$

where α^k and m^k are the vectors of the budget and marginal shares. Inserting into (9.33), multiplying by VC_W and rearranging gives

$$(9.37)\qquad \begin{aligned} z^k\left(\mu - r\iota - \begin{bmatrix} m^k \\ 0 \end{bmatrix}\sigma\right) &= C_W w^k W^k V - \sigma \begin{bmatrix} \alpha^k C^k \\ 0 \end{bmatrix} + C_X\sigma \\ &= V_{\mu C} - C^k\sigma \begin{bmatrix} \alpha^k \\ 0 \end{bmatrix}. \end{aligned}$$

Define P^k_m and P^k_α as the price of a basket of consumption goods that contains exactly one unit of asset g that is consumed in both countries and the expenditure for each asset is $m^k_j P^k_m$ and $\alpha^k_j P^k_\alpha$, respectively. We therewith have in the first case $\frac{m^k_j P_g}{m^k_g P_j}$ units of asset j and in the second case $\frac{\alpha^k_j P_g}{\alpha^k_g P_j}$ units in these baskets.

Let $V_{\mu\alpha^k}$ be the vector of covariances of the asset returns with changes in P^k_α,

$$(9.38)\qquad V_{\mu\alpha^k} = \sigma \begin{bmatrix} m^k \\ 0 \end{bmatrix},$$

and $V_{\mu m^k}$ the vector of covariances of the returns with changes in P_m^k:

$$V_{\mu\alpha^k} = \sigma \begin{bmatrix} \alpha^k \\ 0 \end{bmatrix}. \tag{9.39}$$

Thus we can rewrite (9.37) as

$$z^k \left(\mu - r\iota - V_{\mu m^k}\right) = V_{\mu C} - C^k V_{\mu\alpha^k}. \tag{9.40}$$

We can now aggregate (9.40) over all M domestic investors and get with $z = \sum_{k=1}^M z^k$, $C = \sum_{k=1}^M C^k$, $P_\alpha = \sum_{k=1}^M P_\alpha^k \frac{C^k}{C}$ and $P_m = \sum_{k=1}^M P_m^k \frac{z^k}{z}$:

$$z(\mu - r\iota - V_{\mu m}) = V_{\mu C} - CV_{\mu\alpha}. \tag{9.41}$$

The term $1/P_m$ denotes the real value of a marginal increase in domestic consumption as the result of a change of the value of the portfolio domestic investors hold, i.e. P_m depends on the preferences of the domestic investors.

For a foreign investor we get the equivalent equation in foreign currency:

$$z^*(\mu^* - r^*\iota - V_{\mu^* m^*}) = V_{\mu^* C^*} - C^* V_{\mu^*\alpha^*}. \tag{9.42}$$

In order to aggregate domestic and foreign investors both equations have to be denoted in the same currency. Define $z^F = ez^*$, $C^F = eC^*$, $P_m^F = eP_m^*$ and $P_\alpha^F = eP_\alpha^*$ as the corresponding values in domestic currency. With Ito's lemma we get:

$$V_{\mu^*\alpha^*} = V_{\mu^* 1/e\alpha^F} = V_{\mu^*\alpha^F} - V_{\mu^* e}, \tag{9.43}$$

$$V_{\mu^* m^*} = V_{\mu^* 1/em^F} = V_{\mu^* m^F} - V_{\mu^* e}, \tag{9.44}$$

$$V_{\mu^* C^*} = V_{\mu^* 1/eC^F} = \frac{1}{e}\left(V_{\mu^* C^F} - V_{\mu^* e} C^F\right). \tag{9.45}$$

From (9.22) and (9.23) we see that for the excess returns of the riskless assets of the other country we have

$$\frac{dH_{n+1^*}}{H_{n+1^*}} = -\frac{dH_{n+1}^*}{H_{n+1}^*} + \left(\frac{de}{e}\right)^2 \tag{9.46}$$

and from (9.20) and (9.21) for the risky assets:

$$\frac{dH_i}{H_i} = \frac{dH_i^*}{H_i^*} + \frac{de}{e}\frac{dP_i}{P_i}. \tag{9.47}$$

By defining $L = \begin{bmatrix} I_n & 0 \\ 0 & -1 \end{bmatrix}$ we get from (9.46) and (9.47):

$$L(\mu^* - r^*\iota) = (\mu - r\iota) + V_{\mu^* e}. \tag{9.48}$$

It can further be shown that for any stochastic variable y that

$$LV_{\mu y} = V_{\mu^* y}. \tag{9.49}$$

Multiplying (9.42) with L and inserting (9.43), (9.45), (9.48) and (9.49) we get

$$z^F \left(\mu - r\iota - V_{\mu m^F}\right) = V_{\mu C^F} - C^F V_{\mu \alpha^F}. \tag{9.50}$$

This relation is now expressed in domestic currency and can therefore be aggregated with equation (9.41). By defining $z^W = z + z^F$, $P_m^W = P_m + P_m^F$ and $C^W = C + C^F$ we get

$$z^W \left(\mu - r\iota - V_{\mu m^W}\right) = V_{\mu C^W} - C^W V_{\mu \alpha^W}. \tag{9.51}$$

By denoting $c = \frac{C^W}{P_\alpha^W}$ the real world consumption we get with Ito's lemma:

$$P_\alpha^W V_{\mu c} = P_\alpha^W V_{\mu C^W / P_\alpha^W} = V_{\mu C^W} - C^W V_{\mu \alpha}. \tag{9.52}$$

Inserting into (9.51) and rearranging gives:

$$\mu - r\iota - V_{\mu m^W} = \frac{P_\alpha^W}{z^W} V_{\mu c}. \tag{9.53}$$

This relation has to hold for any asset and any portfolio M (not necessarily the market portfolio):

$$\mu_M - r\iota - V_{\mu_M m^W} = \frac{P_\alpha^W}{z^W} V_{\mu_M c}. \tag{9.54}$$

Solving for the preference parameter $\frac{P_\alpha^W}{z^W}$ and inserting into (9.53) gives the final relationship

$$\mu - r\iota - V_{\mu m^W} = \left(\mu_M - r\iota - V_{\mu_M m^W}\right) \frac{V_{\mu c}}{V_{\mu_M c}}. \tag{9.55}$$

The portfolio M can freely be chosen, a useful approach in line with the traditional CAPM is to choose the world market portfolio of all jointly risky assets and the riskless asset of the other country.[5]

The left hand side of equation (9.55) denotes the real excess return of the asset, on the right side the first term denotes the real excess return of the reference portfolio and the last term denotes the covariance of the nominal asset returns with world real consumption relative to this relation of the reference portfolio. The real expected excess returns are higher the higher this covariance is. A company that produces a product whose demand depends heavily on aggregate consumption would have a higher covariance. The excess returns depend on the product the company produces and the country in which their product is demanded.

This model showed that different tastes across countries do not affect the allocation of assets and their real expected returns, they only depend on the relation to world aggregate consumption, i.e. aggregated tastes. These findings are very similar to those of the CCAPM. By using a portfolio M as reference whose returns are perfectly correlated with changes in aggregate consumption, (9.55) equals the CCAPM in real terms.

To get differences in the allocation of assets, e.g. the observed bias towards domestic assets we have either to deviate from the law of one price for consumption goods and assets, e.g. by introducing transportation costs or to introduce barriers to international investment, what will be done in the next section.

9.3 Barriers for international investment

Stulz (1981b) provides a model of international asset pricing in the presence

[5] Stulz (1981a, p. 392) shows further that the portfolios held by investors in different countries differ only by the share they invest into the others country riskless asset, the relative shares of the jointly risky assets are equal and within a country the investors differ only in the share of the domestic riskless asset. This finding is very close to the Tobin separation theorem of the traditional CAPM.

of barriers to international investment. Like in the last section he assumes two countries, in each country there are n and n^* assets located, such that $n+n^* = N$. We assume that foreign investors are free to invest in their home country or abroad without facing any restrictions, whereas domestic investors can invest into domestic assets without any restrictions, but if they want to invest in foreign asset have to pay a tax of θ per time period and invested unit of wealth. This tax can either be a transaction tax, but it can also be interpreted as costs of obtaining additional information.

The expected return from holding a foreign asset is reduced to $\mu_i - \theta$ and from being short in an asset to $-\mu_i - \theta$. The tax of θ has not to be paid on the net position of a foreign asset, but on every position individually, i.e. holding one unit of a foreign asset long and one unit short results in a tax of 2θ.

With these conditions we assume every investor to optimize his portfolio holdings according to portfolio theory presented in section 4.2. If we denote the covariance matrix of all domestic and foreign assets by V and w^k the long and v^k the vector of short positions of the kth domestic investor the variance of a portfolio is given by

$$V_p = (w^k - v^k)'V(w^k - v^k). \tag{9.56}$$

The return of this portfolio consists of the expected returns from holding the assets, $(w^k - v^k)'\mu$, subtracting the taxes for holding foreign assets, $(w^k + v^k)'\iota_n\theta$, where ι_n denotes a vector with zeros in the first n and ones in the last n^* rows, and the return from holding the risk free asset $(1 - (w^k - v^k)'\iota)r$:[6]

$$\mu_p = (w'k - v^k)'\mu - (w^k - v^k)'\iota_n\theta + (1 - (w^k - v^k)'\iota)r. \tag{9.57}$$

From portfolio theory we know that to find the efficient frontier we have to minimize (9.56) subject to the constraints $\mu_p \geq \mu^0$, $w^k \geq 0$ and $v^k \geq 0$.[7] With λ^k

[6] We assume the exchange rate not to vary in this model, i.e. the asset is riskless in both countries. Instead of considering two different countries it is also possible to interpret this model with two groups of investors facing different costs of investing into certain assets.

[7] For practical reasons we minimize $\frac{1}{2}V_p$ to avoid a 2 in the first order conditions.

denoting the Lagrangean multiplier for the first constraint we get the following first order conditions with L^k denoting the Lagrange function:

$$\frac{\partial L^k}{\partial w^k} = V(w^k - v^k) - \lambda^k(\mu - r\iota - \theta\iota_n) \geq 0, \tag{9.58}$$

$$\frac{\partial L^k}{\partial v^k} = -V(w^k - v^k) + \lambda^k(\mu - r\iota + \theta\iota_n) \geq 0, \tag{9.59}$$

$$(w^k)'\frac{\partial L^k}{\partial w^k} = 0, \tag{9.60}$$

$$(v^k)'\frac{\partial L^k}{\partial v^k} = 0. \tag{9.61}$$

By combining (9.58) and (9.59) we get

$$\lambda^k(\mu - r\iota + \theta\iota_n) \geq V(w^k - v^k) \geq \lambda^k(\mu - r\iota - \theta\iota_n), \tag{9.62}$$

or with V_i denoting the ith column of V, i.e. $V_i(w^k - v^k)$ denoting the covariance of the ith asset with the portfolio $w^k - v^k$, and

$$1_F(i) = \begin{cases} 1 & \text{if the asset is foreign and the investor is domestic} \\ 0 & \text{if the asset is domestic or the investor is foreign} \end{cases} \tag{9.63}$$

we can rewrite (9.62) as

$$\lambda^k(\mu_i - r + \theta 1_F(i)) \geq V_i'(w^k - v^k) \geq \lambda^k(\mu_i - r - \theta 1_F(i)). \tag{9.64}$$

For domestic assets and foreign investors (9.64) reduces to

$$\lambda^k(\mu_j - r) = V_j'(w^k - v^k). \tag{9.65}$$

If the second inequality in (9.64) holds strict we know from (9.58) and (9.60) that

$$\sum_{j=1}^{N} w_j^k \left(V_j'(w^k - v^k) - \lambda^k(\mu_j - r - \theta 1_F(i))\right) = 0$$

holds only if $w_i^k = 0$. Otherwise this sum must be positive by the restriction that $w_i^k \geq 0$ and the second term being positive. Equivalently with (9.61) the strictness of the first inequality in (9.64) implies $v_i^k = 0$. If both inequalities hold strictly the asset is not held by the investor. Hence one of the two inequalities has to be an equality. Defining for a domestic investor

$$\pi^k(i) = \begin{cases} 1 & \text{for a foreign asset held long} \\ 0 & \text{for a domestic asset} \\ -1 & \text{for a foreign asset held short} \end{cases} \tag{9.66}$$

we can write

$$V_i'(w^k - v^k) = \lambda^k(\mu_i - r - \theta\pi_k(i)). \tag{9.67}$$

Dividing (9.67) by (9.65) gives us

$$\frac{V_i'(w^k - v^k)}{V_j'(w^k - v^k)} = \frac{\mu_i - r - \theta\pi_k(i)}{\mu_j - r} \tag{9.68}$$

for $i \neq j$. The N net demands $w^k - v^k$ are the only unknown variables in these $N-1$ equations and the relative demands of the assets can be determined.[8] As no preferences enter equation (9.68) the relative demand for every domestic investor only depends on the properties of the assets and hence all investors have the same relative demand for risky assets, i.e. have the same optimal risky portfolio. But as we will see it has not to be the world market portfolio. The optimal risky portfolio for foreign investors will be different of that for domestic investors unless the term $\theta\pi^k(i)$ equals zero, i.e. no barriers for investment exist. Nevertheless all foreign investors will demand the same optimal risky portfolio.

Suppose that there exists a foreign asset whose return is uncorrelated with any other asset, domestic or foreign, i.e. $\sigma_{ij} = 0$ for $i \neq j$. Suppose further that the excess return is zero, as it is in the traditional CAPM, i.e. $\mu_i - r = 0$. Then (9.64) becomes

$$\lambda^k\theta \geq \sigma_i^2(w_i^k - v_i^k) \geq -\lambda^k\theta. \tag{9.69}$$

For holding the asset long we need $\sigma_i^2 w_i^k = -\lambda^k\theta$. As λ^k can be shown to be positive,[9] this implies $w_i^k < 0$. In the same manner holding the asset short implies $v_i^k < 0$. Both results violate the assumption of nonnegativity of the positions. Therefore an asset with such a property is not traded by domestic investors. This result shows that not all assets are held by domestic investors.

[8] Provided that there are no assets that are perfect substitutes for the other assets, e.g. perfectly correlated assets or mutual funds. Furthermore an asset is either held long or short, the tax avoids holding both positions as only the net position is important for the characteristics of the portfolio.

[9] λ^k is the shadow price of a gain in expected return.

Define now g_i^k and Q_i^k such that (9.64) becomes

$$\lambda^k(\mu_i - r + \theta 1_F(i)) + \lambda^k Q_i^k = V_i'(w^k - v^k) = \lambda^k(\mu_i - r - \theta 1_F(i)) + \lambda^k q_i^k. \tag{9.70}$$

As for domestic assets $1_F(i) = 0$ we find that unless $\lambda^k = 0$, $q_i^k = Q_i^k = 0$ and for foreign assets $q_i^k + Q_i^k = 2\theta$.

Define w^s as the vector of the share the assets have on total world wealth W^W and let with W^k as the wealth of investor k define: $T^k = \lambda^k W^k$, $T^D = \sum_D T^k$, $T^F = \sum_F T^k$, $\delta^k = \frac{T^k}{T^D}$, $\gamma^D = \frac{T^D}{T^D+T^F}$, $q^k = (q_1^k, \ldots, q_N^k)'$ and $q^D = \sum_D \delta^k q^k$, where D and F denote summing over all domestic and foreign investors, respectively.

Multiplying (9.62) by W^k and summarizing over all investors gives

$$V \sum_{D+F} (w^k - v^k) W^k = (T^D + T^F)(\mu - r\iota + \theta\iota_n) + \sum_{D+F} T^k q^k.$$

Inserting the above definitions and using that $\theta\iota_n = q_i^k = 0$ for foreign investors we get

$$\begin{aligned} V w^s W^W &= (T^D + T^F)\left(\mu - r\iota + \theta\iota_n \frac{T^D}{T^D + T^F} \right. \\ &\qquad \left. + \frac{1}{T^D + T^F} \sum_{D+F} T^k q^k \right) \\ &= (T^D + T^F)\left(\mu - r\iota + \gamma^D \theta\iota_n \right. \\ &\qquad \left. + \frac{T^D}{T^D + T^F} \sum_D \frac{T^k}{T^D} q^k \right) \\ &= (T^D + T^F)\left(\mu - r\iota + \gamma^D\theta\iota_n + \gamma^D \sum_D \delta^k q^k \right) \\ &= (T^D + T^F)\left(\mu - r\iota + \gamma^D\theta\iota_n + \gamma^D q^D\right). \end{aligned} \tag{9.71}$$

Multiplying (9.71) by w^s and defining σ_m^2 as the variance of the world market portfolio m, $\theta_m = (w^s)'\theta\iota_n$ as the tax to be paid by a domestic investor for holding the world market portfolio and $q_m = (w^s)'q^D$, we get with μ_m as the expected return on the world market portfolio:

$$\sigma_m^2 W^W = (T^D + T^F)(\mu_m - r - \theta_m \gamma^D + q_m \gamma^D). \tag{9.72}$$

Define $\beta^m = \frac{Vw^s}{\sigma_m^2}$ as the vector of betas of the assets with the world market portfolio. Solving (9.72) for $T^D + T^F$ and inserting into (9.71) we get

$$\mu - r\iota - \gamma^D \theta \iota_n + \gamma^D q^D = \beta^m (\mu_m - r - \theta_m \gamma^D + q_m \gamma^D). \tag{9.73}$$

With $\theta = 0$ we get the usual relationship of the CAPM as derived earlier.[10] For domestic assets this reduces to

$$\mu_i - r = \beta_i^m (\mu_m - r - \theta_m \gamma^D + q_m \gamma^D). \tag{9.74}$$

We find a linear relationship between the beta of the asset with the world market portfolio and the excess return of the asset. the slope can be either smaller or larger than in the CAPM, where it is $\mu_m - r$.[11]

For a foreign asset we get from (9.73):

$$\mu_i - r = \beta^m (\mu_m - r - \theta_m \gamma^D + q_m \gamma^D) + \gamma^D \theta - \gamma^D q_i^D. \tag{9.75}$$

We find a similar Security Market Line as for domestic assets, it has the same slope, but it is shifted. If the asset is hold long, the second inequality in (9.64) becomes an equality and by inspection of (9.70) we see that $q_i^k = 0$ and hence $q_i^D = \sum \pi^k q_i^k = 0$. The SML is shifted by $\gamma^D \theta$ upwards. If the asset is held short it follows in the same way that $Q_i^D = 0$ and hence from $q_i^D + Q_i^D = 2\theta$ that $q_i^D = 2\theta$ and therefore the SML is shifted downwards by $\gamma^D \theta$. We find to have three SML, one for domestic assets, one for foreign assets held long and one for foreign assets held short.

An asset not held by domestic investors fulfills both inequalities in (9.64) strictly, hence we find that for those assets $0 < q_i^k < 2\theta$ and the assets plot between the two SML for foreign assets. Figure 9.1 illustrates these findings.

[10] As $q_i^k = 0$ for domestic and foreign assets if $\theta = 0$ it follows that $q^D = 0$. For domestic assets we always find $q_i^D = 0$.

[11] See Stulz (1981b, p. 930) for a discussion of this topic.

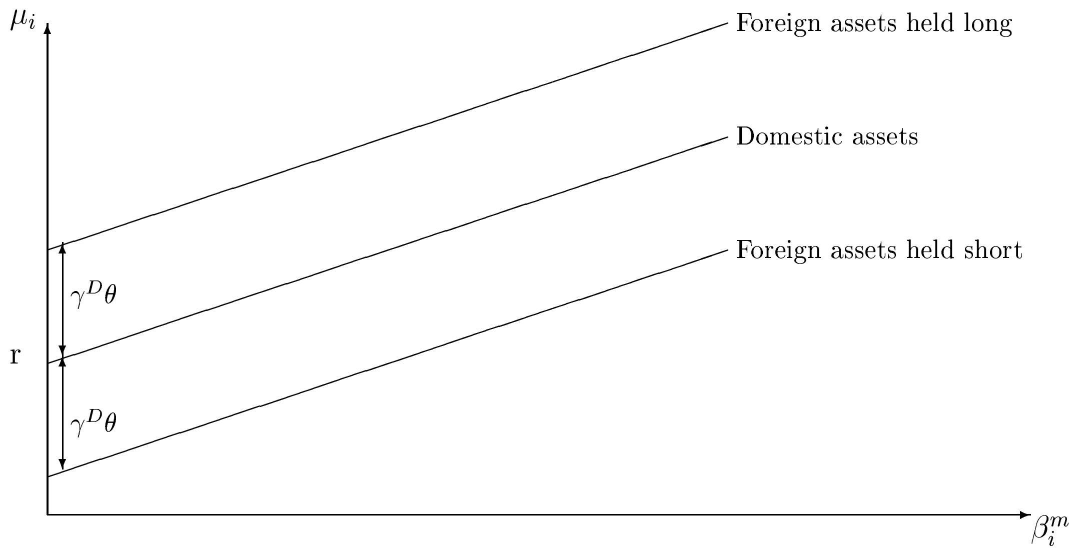

Figure 9.1: The Security Market Lines of the International CAPM

From (9.74) and (9.75) we see that with barriers to international investment an asset with a beta of zero with the world market portfolio may not have the same expected return as the risk free asset if it is a foreign asset. And a beta of one does not ensure to receive the market return.

At last we will show which properties an asset must have that it is held by domestic investors. Aggregating (9.64) over all domestic investors after having multiplied by W^k gives

$$(9.76) \qquad T^D(\mu - r\iota + \theta\iota_n) \geq V\sum_D (w^k - v^k)W^k \geq T^D(\mu - r\iota - \theta\iota_n).$$

Define G^D as the fraction of total wealth invested into risky assets by domestic investors and G^F by foreign investors

$$\begin{aligned} G^D &= \frac{1}{W^D}\sum_{i=1}^N (w_i^k - v_i^k)W^k, \\ G^F &= \frac{1}{W^F}\sum_{i=1}^N (w_i^k - v_i^k)W^k \end{aligned}$$

and $w^D - v^D$ the vectors of fractions invested into each risky asset by domestic investors. The inequalities in (9.76) have to hold strictly for assets not hold by domestic investors. For foreign assets not held we get

$$T^D(\mu_i - r + \theta) > G^D W^D V_i(w^D - v^D) W^k > T^D(\mu_i - r - \theta). \tag{9.77}$$

From (9.65) we similarly get

$$V_i(w^F - v^F) = \frac{T^F}{G^F W^F}(\mu_i - r), \tag{9.78}$$

$V_i^D = V_i(w^D - v^D)$ and $V_i^F = V_i(w^F - v^F)$ denote the covariance of asset i with the portfolio held by domestic and foreign investors, respectively. Solving (9.78) for $\mu_i - r$ and inserting into (9.77) gives us

$$T^D \frac{G^F W^F}{T^F} V_i^F + T^D \theta > G^D W^D V_i^D > T^D \frac{G^F W^F}{T^F} V_i^F - T^D \theta,$$

or after rearranging

$$\theta > \frac{G^D W^D}{T^D} V_i^D - \frac{G^F W^F}{T^F} V_i^F > -\theta. \tag{9.79}$$

Defining $\xi^D = \frac{G^D W^D}{T^D}$ and $\xi^F = \frac{G^F W^F}{T^F}$ enables us to rewrite this as

$$\left| \xi^D V_i^D - \xi^F V_i^F \right| < \theta. \tag{9.80}$$

Foreign assets that have a small covariance with the optimal risky portfolio of the domestic and foreign investors will not be held by domestic investors. These assets then will also have a small covariance and hence a small beta with the world market portfolio, which is a weighted average of these portfolios. the smaller the tax is the more assets will be traded, if $\theta = 0$ all assets are traded as predicted by the CAPM

The reason that assets with a small beta with the world market portfolio are not held is that their expected benefits are too small to overcome the costs. The diversification effect can also be achieved by assigning a higher fraction to the riskless asset in the optimal portfolio.

Also foreign assets that have covariances close to those of domestic assets will not be traded, although their beta may differ significantly from zero. These foreign assets can easily be substituted by domestic assets without imposing the costs. These interpretations show that it is very likely to find domestic investors to hold more domestic assets.

9.4 Empirical evidence[12]

Most empirical investigations test the segmentation or integration of asset markets, i.e they want to find or reject barriers to international investments. In general neither evidence for segmentation nor integration can be found at significant levels. The reason may be that these barriers in most cases affect investors from different countries or different types of investors of the same country not equally. Different tax structures, investment quotas and other factors differ widely between countries and types of investors. This will influence the behavior of asset returns in a much more complicated way than modeled here. Additionally globalization and liberalizations change the rules permanently such that it is difficult to investigate asset returns over time. These difficulties make it very problematic to find support for or against the model.

The only substantial evidence can be derived from assets that differ only in their availability to foreign investors, e.g. the A- and B-shares of the Shenzen Stock Exchange in China. It can only be said that the more widely an asset is available to foreign investors the higher its price is.

[12] Stulz (1995) summarizes the main empirical works on the International CAPM.

10. The Production-based Asset Pricing Model

The CCAPM and the ICAPM presented in the last sections took the productions side of the economy as given and only modeled the demand. Empirically strong evidence has been found that stock returns forecast GDP growth very well. By inverting this relation high expected GDP growth in the future should result in high asset returns now. By taking the demand side (consumption) as given and modeling the supply side (production) of the economy COCHRANE (1991) developed the *Production-based Asset Pricing Model* (*PAPM*).

10.1 The model[1]

We assume a discrete time setting where a single good is produced in a finite number of states.[2] We further have a single asset which pays a dividend in every period depending on the state. The discount factor also depends on the state.

From section 4.1 we know that the fundamental value of the asset with the current state as the only source of information is given by

$$P_t = E\left[\sum_{\tau=t}^{\infty}\left(\prod_{k=t}^{\tau}\rho_s^k\right)D_\tau \,\middle|\, s^t\right], \tag{10.1}$$

where P_t denotes the price at time t, ρ_s^k the discount factor of investor k in state

[1] The derivation follows COCHRANE (1991).

[2] A generalization to a continuous time model, many goods produced and infinite number of states can easily derived, obtaining the same results. See COCHRANE (1991, p. 213).

s^t, D_τ the dividend and s^t the current state. Defining the return as usual we get

$$R_{t+1} = \frac{P_{t+1} + D_{t+1} - P_t}{P_t}. \tag{10.2}$$

Using (10.2) it can easily be verified that

$$P_t = E\left[\rho^s_{t+1}(1 + R_{t+1})P_t | s^t\right]. \tag{10.3}$$

By inserting (10.2) into (10.3) and iterating by inserting (10.3) again, we get

$$\begin{aligned} P_t &= E\left[\rho^s_{t+1}(1 + R_{t+1})P_t | s^t\right] \\ &= E\left[\rho^s_{t+1}(P_{t+1} + D_{t+1}) | s^t\right] \\ &= E\left[\rho^s_{t+1}(\rho^s_{t+2}(1 + R_{t+2})P_{t+1} + D_{t+1}) | s^t\right] \\ &= E\left[\rho^s_{t+1}(\rho^s_{t+2}(P_{t+2} + D_{t+2}) + D_{t+1}) | s^t\right] \\ &= \dots \\ &= E\left[\sum_{\tau=t}^{\infty}\left(\prod_{k=t}^{\tau}\rho^k_s\right) D_\tau \middle| s^t\right], \end{aligned} \tag{10.4}$$

hence (10.3) and (10.1) are equivalent. By dividing (10.3) by P_t we get

$$E\left[\rho^s_{t+1}(1 + R_{t+1}) | s^t\right] = 1. \tag{10.5}$$

For later convenience we rewrite ρ^s_{t+1} in terms of the consumption good. Let $P_0^C(s^{t+1})$ denote the price for the consumption good at time 0 for delivery at time $t+1$ if a certain state s^{t+1} occurs (contingent contract). We assume that such a price exists for every state and date, i.e. we assume a complete market. The real price of this claim in terms of the good at time t in state s^t is given by

$$p^C(s^{t+1}) = \frac{P_0^C(s^{t+1})}{P_0^C(s^t)}. \tag{10.6}$$

We also can express $P_0^C(s^{t+1})$ if we are in a certain state s^t by

$$P_0^C(s^{t+1}) = P_0^C(s^t)\rho^s_{t+1}\pi(s^{t+1}|s^t), \tag{10.7}$$

where $\pi(s^{t+1}|s^t)$ denotes the probability that state s^{t+1} occurs given state s^t. Rearranging and using (10.6) gives

$$\rho^s_{t+1} = \frac{p^C(s^{t+1}}{\pi(s^{t+1}|s^t)}. \tag{10.8}$$

We now turn our attention to the production of the good. The total production, y_t, consists of the production of a single consumption good, c_t, and investments I_t:

$$y_t \equiv c_t + I_t. \tag{10.9}$$

The total production is assumed to follow a certain production function

$$y_t = f(k_t, l_t, s^t), \tag{10.10}$$

where k_t denotes the capital stock and l_t the labor input, which we assume to be constant over time. The increase in the capital stock follows

$$k_{t+1} = g(k_t, I_t), \tag{10.11}$$

where in general $g(k_t, I_t) \neq k_t + I_t$ as we also account for adjustment costs of new investments, i.e. costs like education of employees.

Differentiating (10.10) and (10.11) totally with subscripts denoting the derivatives we have:

$$\begin{aligned}
dk_{t+1} &= g_I(t) dI_t, && (10.12)\\
dk_{t+2} &= g_k(t+1) dk_{t+1} + g_I(t+1) dI_{t+1}, && (10.13)\\
dy_{t+1} &= f_k(t+1) dk_{t+1}. && (10.14)
\end{aligned}$$

We can view the capital stock as fixed for a given period and labor input is fixed over the entire time by assumption, hence the marginal product of capital depends only on the production of the consumption good (sales) as the costs are fixed. It further depends on the adjustment costs of new investments in this period. This marginal product of capital has to equal the return on investment by standard neoclassical theory:

$$1 + R'_{t+1} = \frac{\partial c_{t+1}}{\partial k_{t+1}} g_I(t). \tag{10.15}$$

We assume now that the capital stock for the periods $t+2$ and following is fixed and that the firm has to decide which amount to invest in period t and which amount to invest in period $t+1$.[3] This implies the restriction

$$dk_{t+2} = 0. \tag{10.16}$$

By inserting (10.13) and rearranging we get

$$dI_{t+1} = -\frac{q_k(t+1)}{q_I(t+1)} dk_{t+1}. \tag{10.17}$$

By using (10.9),(10.12),(10.14) and (10.17) we get from (10.15)

$$\begin{aligned} 1 + R'_{t+1} &= \frac{\partial c_{t+1}}{\partial k_{t+1}} g_I(t) \\ &= \frac{dy_{t+1} - dI_{t+1}}{g_I(t) dI_t} g_I(t) \\ &= \frac{f_k(t+1) g_I(t) dI_t + \frac{g_k(t+1)}{g_I(t+1)} g_I(t) dI_t}{dI_t} \\ &= \left(f_k(t+1) + \frac{g_k(t+1)}{g_I(t+1)} \right) g_I(t). \end{aligned} \tag{10.18}$$

If we assume the company to maximize the expected profits by choosing the appropriate investment strategy, standard neoclassical theory suggests that marginal costs and benefits have to equal. By (10.10) the total production in a period is given by the capital stock as the only variable. Total production can only be divided between the production of the consumption good and new investments. An increase in investment by dI_t has to be accompanied by a decrease in sales of consumption goods with the same amount, hence marginal costs of increasing investments are $P^C(s^t) dI_t$. The marginal benefits are the increased production of the consumption good in the next period, due to a rise in total production and a reduction in future investment (the capital at $t+2$ is fixed to a certain amount), i.e. the marginal benefits are $E\left[\frac{P^C(s^{t+1})}{\pi(s^{t+1}|s^t)} dc_{t+1} | s^t\right]$. So the first order condition for a profit maximum is

$$P^C(s^t) dI_t = E\left[\frac{P^C(s^{t+1})}{\pi(s^{t+1}|s^t)} dc_{t+1} | s^t\right]. \tag{10.19}$$

[3] The results have to be valid for only two periods, hence the restriction to periods t and $t+1$ does not change the result.

Inserting (10.9), (10.14) and (10.17) gives

$$P^C(s^t)dI_t = E\left[\frac{P^C(s^{t+1})}{\pi(s^{t+1}|s^t)}\left(f_k(t+1)+\frac{g_k(t+1)}{g_I(t+1)}\right)g_I(t)dI_t|s^t\right]. \tag{10.20}$$

Dividing by $P^C(s^t)dI_t$ and using (10.18), (10.6) and (10.8) we get

$$E\left[\rho_{t+1}(1+R_{t+1})|s^t\right] = 1. \tag{10.21}$$

By comparing this first order condition for a profit maximum on the return on investment with the asset pricing condition (10.5) we see that they both look alike. By writing these conditions in another form we get

$$\sum_{s^{t+1}}\pi(s^{t+1}|s^t)\rho_{t+1}(1+R'_{t+1}) = \sum_{s^{t+1}}\pi(s^{t+1}|s^t)\rho_{t+1}(1+R_{t+1}) = 1. \tag{10.22}$$

It is obvious that the return on investment is a linear combination of asset returns and vice versa. In order to prevent arbitrage, i.e. the company may assets sell short and invest the proceedings to obtain a riskless profit or vice versa, the two expected returns must equal:

$$E\left[R'_{t+1}\right] = E\left[R_{t+1}\right]. \tag{10.23}$$

Having shown the equality of expected asset returns and return on investment, we now can relate the asset returns easily to the growth rate of GDP, i.e. the business cycle. High expected growth of GDP will increase the return on investment and hence expected asset returns. The high expected returns for the near future will lead asset prices to increase as shown in section 4.1. This result will give the result of the stock market leading the business cycle, but this lead should not too far as the precision of expectation decrease with the forecast horizon.

We further have the rationale for asset pricing that the expected asset return should equal the return on investment.

10.2 Empirical evidence

An empirical investigation into the performance of the PAPM can use equation (10.23). Critical to the investigation is the determination of the return on investment, the return on assets can easily be derived from the market. This return can only be determined on the basis of the balance sheet, which is only published quarterly, in Europe often only semiannual or annual, while stock returns are available daily. Further the data in the balance sheet may be biased due to poor accounting standards and actions taken for fiscal measures, i.e. the data quality may only be poor. We therefore face similar problems as in the CCAPM.

The investigation undertaken by Cochrane (1991, pp. 223 ff.) using quarterly data for the period 1947-1987 show nevertheless a high and significant correlation between asset returns and returns on investment. The returns on investment are more volatile than asset returns, suggesting that the arbitrage is not complete due to transaction costs or market incompleteness.

Cochrane (1991, pp. 232 ff.) further finds that the asset returns forecast GDP growth by about 9 months, a not too long period of time as predicted by theory.

These results suggest that in the longer run (for period longer than a quarter) the stock market is driven by the business cycle, where for investigating single assets or industries the business cycle has to be decomposed into the parts relevant for this investigated assets. As other variables of interest to the expected returns, e.g. interest rate or consumption in the CCAPM, are also closly linked to the business cycle, its importance for explaining asset returns is further increased.

What remains an unsolved problem is the explanation of asset returns in the short run. Here only the Conditional CAPM, especially the ARCH specifications, give promising hints.

11. The Influence of Speculation on Asset Prices

The preceding sections offered various models to determine the fundamental value or appropriate returns for assets. Empirical investigations showed support for these models in the long run, but with the exception of the Conditional CAPM clearly failed explaining short term price movements. It was further found that prices can deviate far from the prices predicted by these models, mostly they are too high. These deviations attracted increased attention in recent years, especially in the aftermath of the stock market crash in 1987.

Many models explaining deviations from the fundamental value assume irrational behavior of investors, like fashions, fads or herding.[1] Another group of models emphasizes the importance of imperfect aggregation of information in prices.[2] In this section we will only focus on models assuming rational behavior of all investors.

Deviations from the fundamental value are usually called *bubbles*, if they are the result of rational behavior, they are also called *rational bubbles*.

11.1 Rational bubbles

Like in the static models of asset pricing we assume that investors buy an asset in one period and liquidate it in the following period consuming the proceedings.

[1] See Shiller (1989, pp. 7-68) for an overview.
[2] See Kleidon (1995) for an overview.

In section 4.1 we derived the price of an asset to be the present value of the expected future dividends, i.e. the fundamental value by using the assumption

$$\lim_{k\to\infty}\left(\prod_{j=1}^{n}\frac{1}{1+R_{t+j+1}}P_{t+k}\right)=0. \tag{11.1}$$

If we lift this assumption the fundamental value p_t^* as derived in section 4.1 is only the particular solution to the differential equation

$$P_t=\frac{P_{t+1}+D_{t+1}}{1+R_{t+1}}. \tag{11.2}$$

The general solution to this problem is

$$P_t=P_t^*+B_t, \tag{11.3}$$

where B_t is the bubble term.[3] Inserting (11.3) into (11.2) and noting that the fundamental value as a particular solution also has to fulfill (11.2) we get

$$\begin{aligned} P_t^*+B_t &= \frac{P_{t+1}^*+B_{t+1}+D_{t+1}}{1+R_{t+1}} \\ &= P_t^*+\frac{B_{t+1}}{1+R_{t+1}}. \end{aligned}$$

Rearranging gives the condition for the bubble term:

$$B_{t+1}=(1+R_{t+1})B_t. \tag{11.4}$$

Introducing expectations as in section 4.1.1 equation (11.4) becomes

$$E\left[B_{t+1}|\Omega_t\right]=(1+E\left[R_{t+1}|\Omega_t\right])B_t. \tag{11.5}$$

There are many specifications of the bubble term that satisfy (11.5). The *deterministic bubble*

$$B_{t+1}=(1+E\left[R_{t+1}|\Omega_t\right])B_t \tag{11.6}$$

grows with the same rate as the fundamental value. As long as the expected returns of the fundamental value are positive the bubble term will grow and the

[3] See ASCHINGER (1995, p.120).

price will deviate more and more from the fundamental value.[4] With a negative expected return the bubble term would reduce again. Such a specification of a bubble is unrealistic as empirical evidence shows that in the long run the price follows the fundamental value, i.e. the bubble bursts. The most famous specification that allows the bubble to burst is the *stochastic bubble* as introduced by BLANCHARD/WATSON (1982). They define the bubble as

$$(11.7) \qquad B_{t+1} = \begin{cases} \frac{1+E[R_{t+1}|\Omega_t]}{\pi} B_t + \varepsilon_{t+1} & \text{with probability } \pi \\ \varepsilon_{t+1} & \text{with probability } 1-\pi \end{cases},$$

with $E[\varepsilon_{t+1}|\Omega_t] = 0$ and $Var[\varepsilon_{t+1}|\Omega_t] \geq 0$. With this specification the bubble will burst at some time and the price jumps back near to his fundamental value. To compensate for the the risk of this price jump the expected return is larger as long as the bubble persists and the higher the probability of its burst is.[5] Such a bubble is rational because a further increase in the bubble is justified by the expectation of a further increase with probability π.

This specification of a bubble faces several problems. If we only take assets with limited liability, as shares, into account, a negative price is not reasonable, This allows not for large negative bubble terms.[6] We further see that the burst of a bubble implies a large drop (or in the case of a negative bubble increase) in the price, an event not frequently found in asset markets, where price falls happen to occur over a longer period. On the other hand made this property stochastic bubbles popular for explaining a crash, like the ones of 1929 and 1987,[7] whereas for regular market conditions it is not reasonable.

Other specifications of bubbles that sometimes depend on the fundamental value have also been proposed, but most specifications face the same problems.[8]

Rational bubbles can explain significant deviations of the price from the funda-

---

[4] See ASCHINGER (1995, p. 121).
[5] See ASCHINGER (1995, p. 121).
[6] See CAMPBELL ET AL. (1997, pp. 259 ff.).
[7] See e.g. ASCHINGER (1995, pp. 128 ff.).
[8] See ASCHINGER (1995, pp. 124 f.) and KLEIDON (1995, pp. 481 ff.) for examples and an overview of the literature.

mental value and higher returns than expected from observing the fundamental value, but with the drop of prices if a bubble bursts they are not able to explain price movements in regular markets. They further fail to explain the higher volatility observed in asset markets compared to the fundamental value, the increase in volatility from the presence of bubbles is not sufficient.

Flood/Hodrick (1990, p. 87) point out that it is empirically very difficult to proof the existence of bubbles. Critical is the determination of the fundamental value and the exact specification of the model. Therefore empirical evidence for bubbles are rarely found.[9]

11.2 Rational speculation

Leach (1991) presented a model similar to the rational bubbles in the specification of Blanchard/Watson (1982), but he allows the investors to have time horizons of more than two periods. We have S types of investors, each having a time horizon of $2 < J < \infty$ periods.

We assume that there exists only a single unit of a single asset owned by a single investor. Every period a new investor enters the market and the owner of the asset offers him the asset at a fixed price. The entrant can accept the offer and become the new owner or he can reject and leave the market.

Every investor belongs to one of the S types of investors. Each type s ($s = 1, \ldots, S$) assigns a different value to the asset. This assumption can be justified by assuming different expectations of future dividends. For simplicity the fundamental value is assumed to be constant over time for all types.

The owner receives the dividend at the beginning of a period, then learns which type of investor will enter the market, correctly interferes the entrants reservation price and then makes his offer to sell. He does not know the types of entrants in

[9] See Kleidon (1995, p. 482).

the following periods, it is only known the probability that the entrant will be of type s. This probability is assumed not to vary over time.

If the owner wants to sell the asset and leave the market afterwards he quotes a price that equals the entrants reservation price,[10] which the entrant accepts. If he does not want to sell the asset, he quotes a higher price than the reservation price of the entrant and the entrant rejects the offer.[11]

The owner has to decide every period whether to sell the asset at the entrants reservation price or to hold the asset for another period. At the end of his life, in period J, he has to sell the asset to any entrant. If the entrants reservation prices is low and the owner is young, he will not sell the asset, even if the price is above what he thinks is his fundamental value, because he *rationally speculates*[12] that at a later date he may receive a higher price. An old investor on the other hand may be willing to sell the asset even below his fundamental value fearing to receive an even lower price later.[13]

Therefore we observe prices that may differ from the fundamental value of the owner to both sides. But as the owner will in general not hold the asset until he is old, the deviations will more frequently be on the upper side unless the owner has not a very high fundamental price.

The prices will vary around the fundamental values of the owners, i.e. the volatility of the asset prices will be higher than the volatility of any fundamental value of the owner. Hence this model is able to explain short-term deviations from the fundamental value, whereas in the long run prices will follow the fundamental

[10] To quote a lower price would violate the assumption of rationality as larger profits would be possible. The reservation price of the entrant has not necessarily to be his fundamental value. Although the entrant may assign a low fundamental value to the asset he may be willing to enter the market, speculating to sell the asset at an even higher price.

[11] Alternatively to this mechanism an auction can be conducted every period giving the same results.

[12] Aschinger (1995, p. 17 f.) defines *speculation* in a narrow sense as buying (or in our sense holding) or selling a good in order to receive a profit from a later selling or buying, not to make profits from holding the asset (e.g. from dividends).

[13] Leach (1991) provides formal proofs of these statements together with proving the existence and uniqueness of the equilibrium.

value.

Empirically this model is even harder to confirm than the rational bubbles. Beside the problem of the specification of the model, we have to determine not only one, but S different fundamental values. On the other hand we should expect a higher volatility of prices compared to an "average" fundamental value. As this empirically can be found,[14] this model may provide a platform to explain these findings. A direct test of this model has thus far not been proposed in the literature.

[14] See e.g. LeRoy (1990).

12. The Theory of Financial Disequilibria

The models described in the preceding sections of this chapter all derived expressions for *equilibrium* prices or expected returns of assets. Different variables affecting this equilibrium and the mechanism through which they affect the equilibrium have been proposed. These variables all affect the fundamental value of the assets are therefore also called *real factors*.

Short-term deviations from the fundamental values may be explained by rational speculation. These deviations also form an equilibrium. Empirical investigations suggest that deviations from the fundamental value do not only occur in the short run, but can last for months, even for years. The most cited examples are the long Hausse from 1982 to October 1987 and the current Hausse lasting since 1991. These long lasting deviations usually do not affect only a certain asset or industry, but the market as a whole, i.e. instead of relative prices the price level is affected.[1]

Market commentators, unable to find a change in the fundamental values, often explain larger increases of the market prices with liquidity pouring into the market and larger falls with liquidity flowing out of the market. These movements are often, but not necessarily, accompanied by an adverse movement in bond prices. These observations suggest that the *flow of funds* has another important role in determining the asset price level, rather than relative prices of assets.

If the economy were in equilibrium no excess liquidity would be present to in-

[1] See PEPPER (1994, p. 9).

fluence asset prices. Therefore PEPPER (1994) emphasizes the importance of *financial disequilibria* to explain the price level of assets.

12.1 Financial Disequilibrium

In reality the economy is unlikely to be in equlibrium. PEPPER (1994, ch. 15) illustrates this with the following scenario:

Assume the nominal GDP grows by a relative large amount. This can be due to a growth of real activity or a high inflation. In both cases the companies have an increased demand for finance, they either need more credits or have to issue more shares. Let us concentrate on the demand for bank loans, the effect of issuing new shares is equivalent, also the form of credit, bank loans or bond issues are of no importance. The demand increases because in the first case larger inventories and new investments have to be financed and in the later case the inventories have to be replaced at higher nominal costs.

The supply of credits (savings) does not necessarily increase by the same amount. In case of a growth of real GDP savings may reduce as the confidence in a strong economy grows and more is spent. With high inflation savings may be reduced at first to bring forward expenditure to avoid higher future prices, after these expenditures have been made savings increase, maybe faster than the demand for bank loans. Hence we are likely to have an imbalance in demand and supply of credit. Assume for the following a gap in supply.

This gap will in part be filled by rising interest rates, inducing higher savings and lower demand for credit and in part by commercial banks, maybe with the support of the central bank.[2] Banks do not only act as intermediaires between demand and supply of credit, but also as buffer if they do not exactly match. By

[2] Globalization made it possible to close this gap with international investors. Again the effect does not change. See PEPPER (1994, p. 111).

their ability to create secondary money they will increase the money supply[3] to meet the demand.

Money is held to facilitate transactions (payments) and for savings. The demand for money from transactions is determined by the income of an individual or by the economic activity as a whole. The demand for savings has not necessarily equal the increased supply. Suppose that the remaining supply exceeds the demand. This excess money supply can be

- **unintended savings:** These savings will be spent for consumption, resulting in product-price inflation.
- **intended savings:** If the savers are not content with their interest rate they receive on bank accounts, they will change to other assets, e.g. stocks, if they seem to be more attractive. This increases the price level in these markets. These transactions in the asset markets are not based on information and the assets are not systematically chosen, most assets are affected. These trades are called *liquidity trades*.

In an analogous way excess demand in a recession can be treated. Excess demand and supply of credit and hence flow in and out of the market vary with the business cycle, leading about 6 month.[4]

The real factors varying in the same way with the business cycle should affect asset prices only slightly, smoothing the business cycle by taking into account future developments in the next years. The flow of funds on the other hand swings with the business cycle increasing the movements of the price level as observed in empirical investigations.[5]

We should therefore expect the asset market as a whole to vary with the flow of funds, whereas relative prices of assets should mainly be influenced by real

[3] Money defined in a broad sense, e.g. as M3.
[4] See Pepper (1994, ch. 3).
[5] See Pepper (1994, pp. 58 ff.).

factors.

12.2 Empirical evidence

The investigation of Pepper (1994) shows significant evidence that the flow of funds accounts for a considerable part of the variance in UK stock markets for the period 1967-1989. For time horizons of 1 to 4 years they account for about a third of the variance, whereas real factors only have significant influence for time horizons of more than two years, reaching also about a third for time horizons of four years. For short time horizons of less than a year nor the flow of funds nor the real factors, like in the other asset pricing models, can explain the variation of asset returns. For long time horizons the real factors dominate, because over time excess demand and supply for money cancel and real factors become more visible.

It further turns out that real factors influence relative prices, whereas the flow of funds has no important influence, as predicted by theory.

Pepper (1994, chs. 11-13) further reports a close relation between the development of the stock market and excess money supply for the period 1950-1987 for both the US and UK. The long rise in the stock market from 1982 to the crash of 1987 can be explained with this factor, but the crash itself cannot be explained.

It has to be noted that changes in the concept of monetary policy of the central banks or the regulatory framework for banks can distort the relation between excess money supply and stock prices for a period of adjustment. The funds also often need a change in the real factors to flow in or out of the stock market as an initial condition. But then the flow of funds often disregards changes in real factors, especially changes that should lead to a reversal of the flow, while changes supporting the current flow are viewed as a confirmation and often increase the flow in the same direction.

The theory of financial disequlibrium can explain the long-run deviations from the fundamental value by the flow of funds. It allows also the prices to follow their fundamental values in the long run and explains why markets as a whole deviate from their fundamental values and not only single assets alone as would be suggested by rational bubbles or rational speculation. For the explanation of short term deviations and the high short term volatility this theory gives no hints.

13. Summary

This chapter gave an overview of the main models used in asset pricing, including some more recent developments. Their aim is to determine the fundamental value and/or an appropriate expected return. Most models relate expected returns to risks investors have to bear and have to be compensated for. They differ mainly in the risk factors they allow to enter into the model.

Although a large number of risk factors have been proposed in the literature and many models are used, none of the presented models or any other models used by practitioners are able to explain the observed asset prices or returns sufficiently. While most models work quite well for time horizons of more than a year, they fail in explaining short term movements. Only the Conditional CAPM, especially if covariances are modeled with a GARCH process, shows a satisfactory fit with the data, but the GARCH specification misses any economic reasoning.

The existence of many anomalies[1] observed in asset markets, e.g. seasonal patterns, cannot be explained by any of these theories. They are subject to a large field of theories trying to explain a certain effect or a group of effects. To explain anomalies often even the assumption of rational acting investors, a central element in economic theories, is questioned in these efforts.

As there exists no generally accepted model of asset pricing the opinions of investors concerning the value of an asset diverge substantially. Access to *relevant* information and a superior model are central elements to make extraordinary

[1] Anomalies are returning patterns in asset prices that cannot be explained by any theory and are known to market participants. For an overview of recorded anomalies see LeRoy (1990) or Hawawini/Keim (1995).

profits in the market. Therefore not only academic research but also securities companies make huge efforts to improve the models.[2]

[2] How important it can be to have a superior model can be illustrated with the case of the rise of Salomon Brothers, Inc. In the 1980's due to their ability to value mortgage bonds more precisely than their competitors and gained an outstanding position in this market. When this competitive advantage eroded, the outstanding position of Salomon Brothers, Inc. also declined. See Lewis (1989).

Bibliography

AHN, C. M./CHO, D. C.: Time Varying Risk Preference and Consumption Asset Pricing Model. in Andrew H. Chen (ed.): Research in Finance, Vol. 9, 1991.

ASCHINGER, G.: General Remarks on Portfolio Theory and its Application. Working Paper Nr. 158, Insitute for Economic and Social Sciences, University of Fribourg (Switzerland), 1990.

ASCHINGER, G.: Börsenkrach und Spekulation. München, 1995.

BEEKMANN, W.: Analysis I. Script for Course 01132 at the FernUniversität Hagen, Germany, 1995.

BLACK, F.: Capital Market Equilibrium with Restricted Borrowing. In: Journal of Business, vol. 45, 1972, 444–455.

BLANCHARD, O. J./WATSON, M. W.: Bubbles, Rational Expectations and Financial Markets. Harvard Institute of Economic Research Discussion Paper 877, 1982.

BOEMLE, M.: Unternehmensfinanzierung. 10th ed. edition, Zürich, 1993.

BOLLERLEV, T./CHOU, R. Y./KRONER, K. F.: ARCH modeling in finance: A review of the theory and empirical evidence. In: Journal of Econometrics, vol. 52, 1992, 5–59.

BOLLERSLEV, T.: Generalized Autoregressive Conditional Heteroskedasticity. In: Journal of Econometrics, vol. 31, 1986, 307–327.

BRACHINGER, H.-W./OST, F.: Modelle mit latenten Variablen: Faktorenanalyse, Latent-Structure-Analyse und LISREL-Analyse. in L. Fahrmeier, A. Hamerle, G. Tutz (eds.): Multivariate statistische Verfahren, 1996.

BRACHINGER, H.-W./WEBER, M.: Risk as a Primitive: A Survey of Measures of Perceived Risk. mimeo, University of Fribourg (Switzerland), 1996.

BREEDEN, D. T.: An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities. In: Journal of Financial Economics, vol. 7, 1979, 265–296.

BREEDEN, D. T./GIBBONS, M. R./LITZENBERGER, R.: Empirical Tests of the Consumption-Oriented CAPM. In: Journal of Finance, vol. 44, 1989, 231–262.

BÜSCHGEN, H.: Das kleine Börsen-Lexikon. 19. edition, Düsseldorf, 1991.

CAMPBELL, J. Y./LO, A. W./MACKINLAY, A. C.: The Econometrics of Financial Markets. Princeton, N.J., 1997.

CAMPBELL, J. Y./SHILLER, R. J.: Stock Prices, Earnings, and Expected Dividends. In: Journal of Finance, vol. 53, 1988a, 661–676.

CAMPBELL, J. Y./SHILLER, R. J.: The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. In: Review of Financial Studies, vol. 1, 1988b, 195–228.

COCHRANE, J. H.: Production-Based Asset Pricing and the Link between Stock Returns and Economic Fluctuations. In: Journal of Finance, vol. 46, 1991, 209–237.

CONSTANTINIDES, G. M./MALLIARIS, A. G.: Portfolio Theory, chapter 1, 1–30. Vol. 9, in: JARROW ET AL. (1995).

COX, J. C./INGERSOLL, J. E./ROSS, S. A.: An Intertemporal General Equilibrium Model of Asset Prices. In: Econometrica, vol. 53, 1985, 363–384.

DOKKO, Y./EDELSTEIN, R. H.: Inflation Uncertainty and Stock Market Fluctuations: An Intertemporal CAPM Approach. in: Andrew H. Chen (ed.): Research in Finance, Vol. 9, 1991.

DUMAS, B./ALLAZ, B.: Financial Securities: Market Equilibrium and Pricing Methods. London, 1996.

DUNNE, P. G.: Size and Book-to-Market Factors in a Multivariate GARCH-M Asset Pricing Application. mimeo, 1997.

ENGLE, R. F.: Autorgressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. In: Econometrica, vol. 50, 1982, 987–1007.

ENGLE, R. F./KRONER, K. F.: Multivariate Simultaneous Generalized ARCH. In: Econometric Theory, vol. 11, 1995, 122–150.

ENGLE, R. F./LILIEN, D. M./ROBINS, R. P.: Estimating Time Varying Risk Premia in the Term Structure: The ARCH - M Model. In: Econometrica, vol. 55, 1987, 391–407.

FAMA, E. F.: Efficient Capital Markets: A Review of Theory and Empirical Work. In: Journal of Finance, vol. 25, 1970, 383–423.

FAMA, E. F.: Multifactor Portfolio Efficiency and Multifactor Asset Pricing. In: Journal of Financial and Quantitative Analysis, vol. 31, 1996, 441–465.

FAMA, E. F.: Determining the Number of Priced State Variables in the ICAPM. In: Journal of Financial and Quantitative Analysis, vol. 33, 1998, 217–230.

FAMA, E. F./FRENCH, K. R.: The cross-section of Expected Stock Returns. In: Journal of Finance, vol. 48, 1992, 427–465.

FAMA, E. F./FRENCH, K. R.: Common Risk Factors in the Returns on Stocks and Bonds. In: Journal of Financial Economics, vol. 33, 1993, 3–56.

FAMA, E. F./FRENCH, K. R.: Size and Book-to-Market Factors in Earnings and Returns. In: Journal of Finance, vol. 50, 1995, 131–156.

FAMA, E. F./FRENCH, K. R.: Multifactor Explanations of Asset Pricing Anomalies. In: Journal of Finance, vol. 51, 1996, 55–84.

FAMA, E. F./MACBETH, J. D.: Risk, Return, and Equilibrium: Empirical Tests. In: Journal of Political Economy, vol. 81, 1973, 607–636.

FERSON, W. E.: Theory and Empirical testing of Asset Pricing Models, chapter 5, 145–200. Vol. 9, in: JARROW ET AL. (1995).

FERSON, W. E./HARVEY, C. R.: The Variation of Economic Risk Premiums. In: Journal of Political Economy, vol. 99, 1991, 385–415.

FERSON, W. E./HARVEY, C. R.: Conditioning Variables and the Cross-section of Stock Returns. mimeo, Duke University, 1996.

FLOOD, R. P./HODRICK, R. J.: On Testing for Speculative Bubbles. In: Journal of Economic Perspectives, vol. 4, 1990, 85–101.

GORDON, M. J.: The Investment, Financing, and Valuation of the Corporation. Homewood, Ill., 1962.

HAMILTON, J. D.: Time series Analysis. Princeton, N.J., 1994.

HAWAWINI, G./KEIM, D. B.: On the predictability of Common Stock returns: World Wide Evidence, chapter 17, 497–544. Vol. 9, in: JARROW ET AL. (1995).

INTRILIGATOR, M. D.: Mathematical Optimization and Economic Theory, chapter 13, 326–343. Englewwod Cliffs, N. J., USA, 1971.

JAGANNATHAN, R./WANG, Z.: The Conditional CAPM and the Cross-Section of Expected Returns. In: Journal of Finance, vol. 51, 1996, 3–53.

JARROW, R. A./MAKSIMOVIC, V./ZIEMBA, W. T. eds.: Finance. Handbooks in Operations Research and Management Science, vol. 9. Amsterdam, 1995.

KEYNES, J. M.: The General Theory of Employment, Interest, and Money. Amherst, N.Y., 1997.

KLEIDON, A. W.: Stock Market Crashes, chapter 16, 465–496. Vol. 9, in: JARROW ET AL. (1995).

KNIGHT, F. H.: Risk, Uncertainty and Profit. New York, 1957 (Orig. 1921).

LEACH, J.: Rational Speculation. In: Journal of Political Economy, vol. 99, 1991, 131–144.

LEROY, S. F.: Capital Market Efficiency: An Update. In: Federal Reserve Bank of San Francisco Economic Review, 1, 1990, 29–40.

LEVY, H./SARNAT, M.: Investment and Portfolio Analysis. New York, 1972.

LEWIS, M.: Liar's Poker: Playing the Money Markets. London, 1989.

LINTNER, J.: Security Prices, Risk, and Maximal Gains from Diversification. In: Journal of Finance, vol. 20, 1965a, 587–615.

LINTNER, J.: The Valuation of Risk assets and the selection of Risky Investments in Stock Portfolios and Capital Budgets. In: Review of Economics and Statistics, vol. 47, 1965b, 13–37.

MARKOWITZ, H. M.: Portfolio Selection: Efficient Diversification of Investments. New Haven, Conn. and London, 1959.

MERTON, R. C.: An Intertemporal Capital Asset Pricing Model. In: Econometrica, vol. 41, 1973, 867–887.

MERTON, R. C.: Continuous-Time Finance. Cambridge, Mass. and Oxford, 1990.

NELSON, D. B.: Conditional Heteroskedasticity in Asset Returns: A New Approach. In: Econometrica, 59, 1991, 347–370.

PEPPER, G.: Money, Credit and Asset Prices. London, 1994.

PRATT, J. W.: Risk Aversion in the Small and in the Large. In: Econometrica, vol. 32, 1964, 122–136.

ROLL, R.: A Critique of the Asset Pricing Theory's Test: Part I. In: Journal of Financial Economics, vol. 4, 1977, 129–176.

ROSS, S. A.: The Arbitrage Theory of Capital Asset Pricing. In: Journal of Economic Theory, vol. 13, 1976, 341–360.

SCHNELLER, M. I.: The Arvitrage Pricing Theories: A Synthesis and Critical Review. in: Andrew H. Chen (ed.): Research in Finance, Vol. 8, 1990.

SCHUMANN, J.: Grundzüge der mikroökonomischen Theorie. 6. edition, Berlin, Heidelberg, New York, 1992.

SHARPE, W. F.: Capital Asset Prices: A Theory of Market equilibrium Under Conditions of Risk. In: Journal of Finance, vol. 19, 1964, 425–442.

SHARPE, W. F.: Portfolio Theory and Capital Markets. New York, 1970.

SHILLER, R. J.: Market Volatility. Cambrdge, Mass., 1989.

SMITH, A.: An Inquiry into the Nature and Causes of the Wealth of Nations, Vol. 1. Indianapolis, 1981 (Orig. 1776).

STULZ, R. M.: A Model of International Asset Pricing. In: Journal of Financial Economics, vol. 9, 1981a, 383–406.

STULZ, R. M.: On the Effects of Barriers to International Investment. In: Journal of Finance, vol. 36, 1981b, 923–934.

STULZ, R. M.: International Portfolio Choice and Asset Pricing: An Integrative Survey, chapter 6, 201–223. Vol. 9, in: JARROW ET AL. (1995).

TOBIN, J.: Liquidity Preference as a Behavior Towards Risk. In: Review of Economic Studies, vol. 25, 1958, 65–86.

TOBIN, J.: The Theory of Portfolio Selection. In: in: F. H. Hahn and F. P. R. Brechling: The Theory of Interest Rates, New York, 1966, 3–51.

V. NEUMANN, J./MORGENSTERN, O.: The Theory of Games and Economic Behavior. Princeton, 1953.