MN30380 - Financial Markets

Additional readings

on

Utility theory and portfolio selection

Andreas Krause

Contents

Chapter 1

Utility theory

Since *John von Neumann* and *Oskar Morgenstern* introduced the expected utility hypothesis in 1944, it has become the most popular criterion for modeling decisions under risk. This appendix will give a brief introduction into the reasoning behind expected utility and its implications for risk aversion.

1.1 The expected utility hypothesis

The value, and therewith the returns of assets depend on their future cash flows. These future cash flows usually cannot be predicted with certainty by investors, they are a random variables, hence returns are also random variables. Investment decisions therewith have to be made under risk.[1] In their work von Neumann and Morgenstern (1953) presented a criterion to make an optimal decision if five axioms are fulfilled.[2]

Let $A = \{a_1, \ldots, a_N\}$ be the set of all possible alternatives an individual can choose between,[3] $S = \{s_1, \ldots, s_M\}$ all states that affect the outcome,[4] and $C =$

[1] 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 (1936, p. 68) defines risk as the possibility of the actual outcome to be different from the expected outcome. In contrast, under *uncertainty* the probabilities of each outcome are not known or even not all possible outcomes are known. Cymbalista (1998) provides an approach of asset valuation under uncertainty. In this work we only consider decisions to be made under risk.

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

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

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

$\{c_{11}, \ldots, c_{1M}; \ldots; c_{N1}, \ldots, c_{NM}\}$ 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 $i, j = 1, \ldots, N$ and "$\succeq$" 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 consistently.[5]

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 $a_i = [p_{i1}b_{i1}, \ldots, p_{iM}b_{iM}]$ *and* $b_{ij} = [q_{ij1}c_1, \ldots, q_{ijL}c_L]$. *With* $p^*_{ik} = \sum_{l=1}^{M} p_{il}q_{ilk}$[6] *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 2) and that more complex lotteries can be build up from their basic elements (axiom 3).

[5] 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.).

[6] 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.

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* $a_i \succ b_i$*.*

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 from these axioms are identical to those to be derived in an instant.[7]

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 $a_i = [p_{i1}c_1, \dots, p_{iM}c_M]$, where without loss of generality $c_1 \succ c_2 \succ \cdots \succ c_M$. Such an order is ensured to exist by axiom 1.

Using axiom 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 3 to substitute c_i by c_i^* in a_i:

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

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

This alternative only has two possible outcomes: c_1 and c_M. By applying axiom 2 we get

$$a_i \sim [p_i c_1, (1-p_i)c_M]$$

with $p_i = \sum_{j=1}^{M} u_{ij} p_{ij}$, what is the definition of the expected value for discrete random variables: $p_i = E[u_i]$.[8]

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 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 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 $E[U(a_i)] > E[U(a_j)]$. □

The criterion to choose between two alternatives, 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.

1.2 Risk aversion

> "Individuals are *risk averse* if they always prefer to receive a fixed payment to a random payment of equal expected value."[9]

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

[9] Dumas and 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.[10] The Arrow-Pratt measure is the most widely used concept to measure this risk aversion. We will derive this measure following Pratt (1964), a similar measure has independently also been developed by Arrow (1963).

With the definition of risk aversion above, an individual prefers to receive a fixed payment of $E[x]$ to a random payment of 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 $E[x]-\pi$ and x. By applying the expected utility principle we see that the expected utility of these two payments has to be equal:

$$E\left[U\left(x\right)\right] = E\left[U\left(E\left[x\right]-\pi\right)\right] = U\left(E\left[x\right]-\pi\right). \tag{1.1}$$

The term $E\left[x\right]-\pi$ is also called the *cash equivalent* of x. Approximating the left side by a second order Taylor series expansion around $E\left[x\right]$ we get[11]

$$\begin{aligned} E\left[U\left(x\right)\right] &= E\Big[U\left(E\left[x\right]\right) + U'(E[x])(x-E[x]) \\ &\quad + \frac{1}{2}U''(E[x])(x-E[x])^2\Big] \\ &= U(E[x]) + U'(E[x])E[x-E[x]] \\ &\quad + \frac{1}{2}U''(E[x])E[(x-E[x])^2] \\ &= U(E[x]) + \frac{1}{2}U''(E[x])Var[x]. \end{aligned} \tag{1.2}$$

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

$$U(E[x]-\pi) = U(E[x]) + U'(E[x])\pi. \tag{1.3}$$

[10] Despite this clear evidence for risk aversion, many economic theories assume that individuals are risk neutral. Prominent examples in finance are the information-based models of market making (see section **??**) and several asset pricing theories.

[11] We assume higher order terms to be negligible, what can be justified if x does not vary too much from $E\left[x\right]$.

Inserting (1.2) and (1.3) into (1.1) and solving for the risk premium π we get

$$\pi = \frac{1}{2}\left(-\frac{U''(E[x])}{U'(E[x])}\right) Var(x). \tag{1.4}$$

PRATT (1964) now defines

$$z = -\frac{U''(E[x])}{U'(E[x])} \tag{1.5}$$

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 x, $Var[x]$,[12] hence the other term in (1.4) can be interpreted as risk aversion. Defining $\sigma^2 = Var[x]$ we get by inserting (1.5) into (1.4):

$$\pi = \frac{1}{2} z\sigma^2. \tag{1.6}$$

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. $U'(E[x]) > 0$, then this implies that $U''(E[x]) < 0$. This relation is also known as the first Gossen law and states the saturation effect.[13] The assumption of risk aversion is therefore in line with the standard assumptions in microeconomic theory.

The conditions $U'(E[x]) > 0$ and $U''(E[x]) < 0$ imply a concave utility function. The concavity of the function (radius) is determined by the risk aversion.[14]

Figure 1.1 visualizes this finding for the simple case of two possible outcomes, x_1 and x_2, having equal probability of occurrence.

[12] A justification to use the variance as a measure of risk is given in appendix 2.

[13] See SCHUMANN (1992, p. 49).

[14] 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, the second derivative of the utility function has to be positive, hence it is convex.

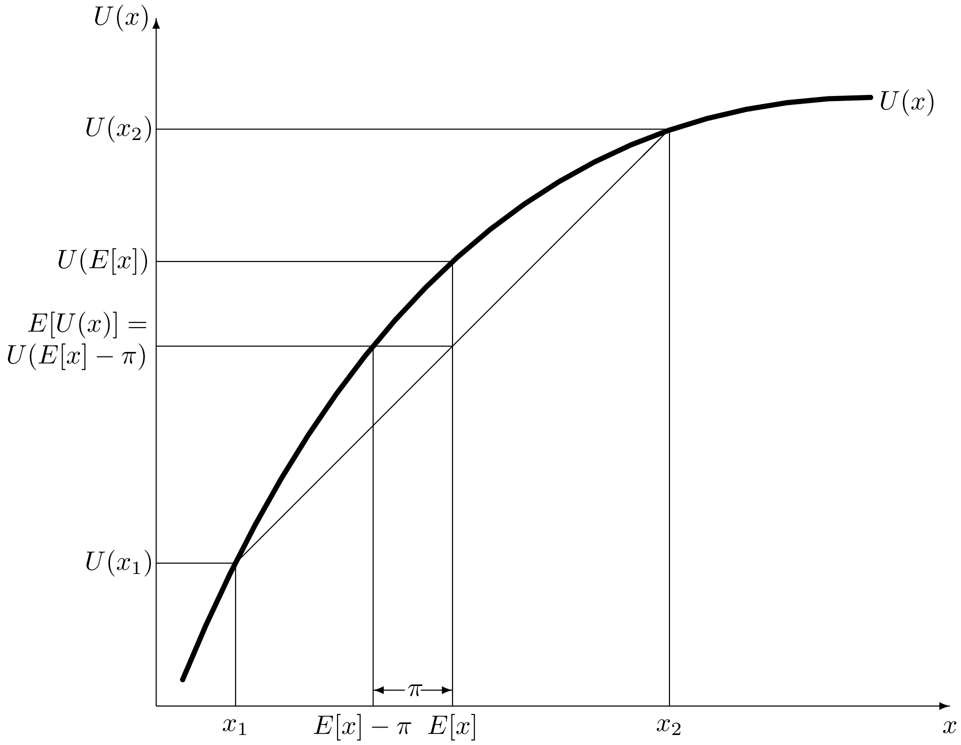

Fig. 1.1: The Arrow-Pratt measure of risk aversion

Chapter 2

The portfolio selection theory

When considering to invest into asset markets, an investor has to make three decisions:

- the amount he wants to invest into the asset market,
- determine the assets he wants to invest in,
- determine the amount he wants to invest into each selected asset.

This appendix describes a method how to make these decisions and find an *optimal portfolio*.[1] 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 contingencies. The investor should build toward an integrated portfolio which best suites his needs."[2]

For this reason the associated theory is called *portfolio selection theory* or short *portfolio theory*, rather than asset selection.[3] 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] A *portfolio* is the entirety of all investments of an individual. See BÜSCHGEN (1991, p.552).
[2] MARKOWITZ (1959, p.3).
[3] See MARKOWITZ (1959, p.3).

2.1 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 a criterion that uses only observable variables instead of individual utility functions. For this purpose many criteria have been proposed,[4] 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.

The *mean-variance criterion* is the most popular criterion not only in finance. The reason is first that it is easy 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-variance criterion, which by this mean 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.[5] The outcome is characterized by its expected value, the mean, and its risk, measured by the variance of outcomes.[6]

The mean-variance criterion is defined as

$$a_i \succeq a_j \Leftrightarrow \begin{cases} Var[a_i] < Var[a_j] \quad \text{and} \quad E[a_i] \geq E[a_j] \\ \text{or} \\ Var[a_i] \leq Var[a_j] \quad \text{and} \quad E[a_i] > E[a_j] \end{cases} . \tag{2.1}$$

It is a 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

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

[5] See Levy and Sarnat (1972, pp. 307 ff.).

[6] 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 and Weber (1997) 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.

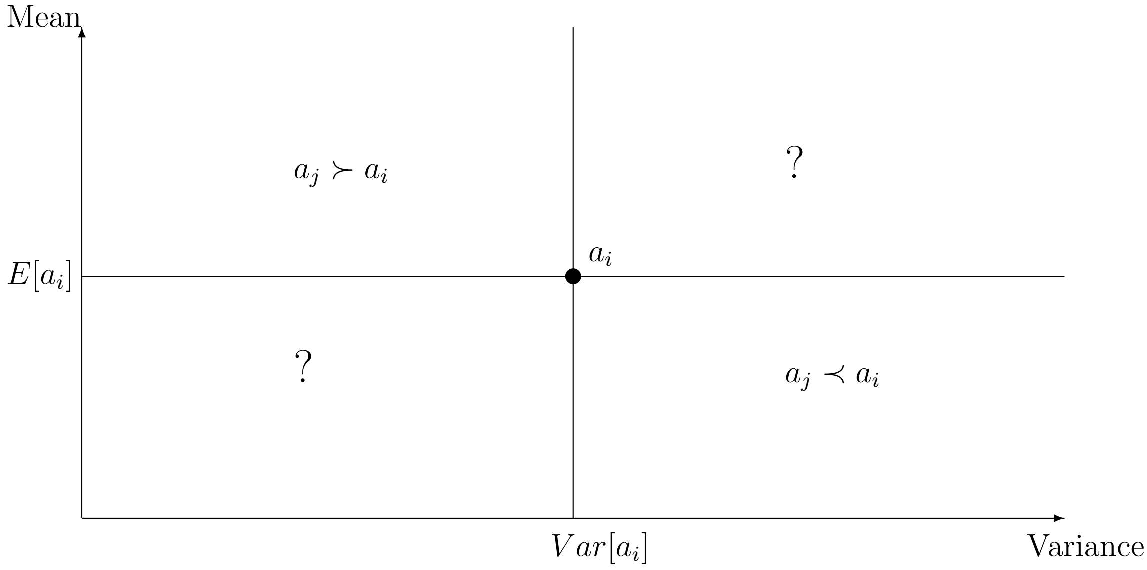

Fig. 2.1: The mean-variance criterion

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.[7]

In figure 2.1 an alternative in the upper left and lower right areas can be compared to a_i by using the mean-variance criterion, while in the areas marked by "?" 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,[8] 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 2.2 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 alternative are called *efficient* and all efficient alternatives form the *efficient frontier*.[9] 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

[7] See Levy and Sarnat (1972, pp. 308).
[8] We will see that this condition is fulfilled in the case of portfolio selection for all relevant portfolios.
[9] See Levy and Sarnat (1972, pp. 318 ff.).

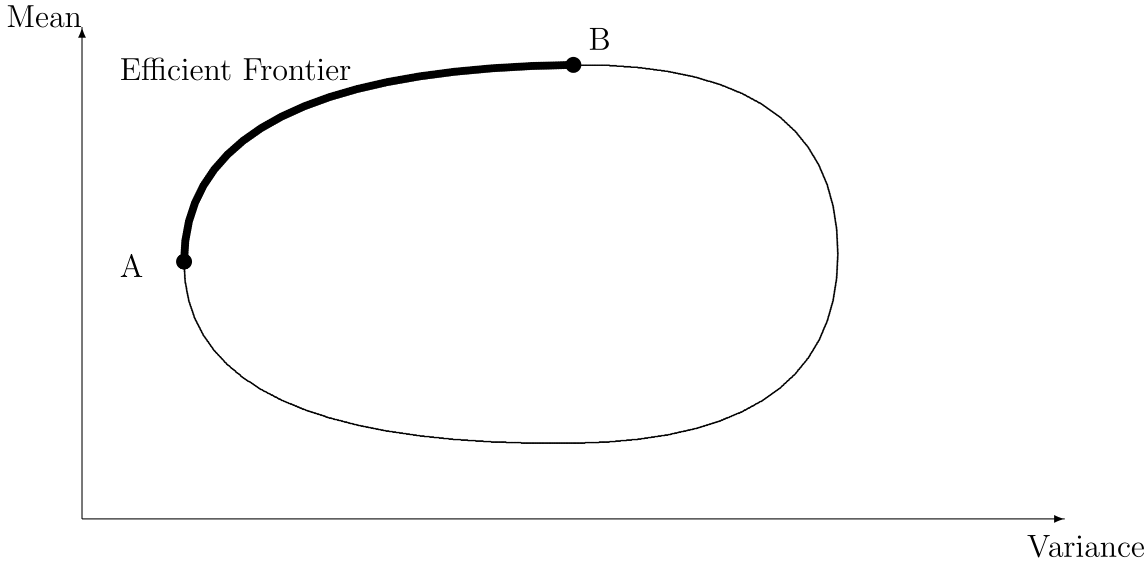

Fig. 2.2: The efficient frontier

true preferences are not always reflected by the results of this criterion.[10] If the utility function is quadratic, we will show that the mean-variance criterion always reflects the true preferences.[11]

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 $b_1 = 1$.[12] The utility function and its derivatives are therefore given by

$$U(x) = x + bx^2, \tag{2.2}$$

$$U'(x) = 1 + 2bx, \tag{2.3}$$

$$U''(x) = 2b. \tag{2.4}$$

According to (1.5) the Arrow-Pratt measure of risk aversion turns out to be

$$z = -\frac{2b}{1 + 2bE[x]}. \tag{2.5}$$

[10] See LEVY AND 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.

[11] See LEVY AND SARNAT (1972, pp. 379 ff.).

[12] 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 $y \to \frac{y-b_0}{b_1}$ to achieve the normalization. See LEVY AND SARNAT (1972, pp. 205 and 379).

If we concentrate on risk averse individuals and assume reasonably positive marginal utility, (2.5) implies that

$$b < 0. \tag{2.6}$$

But if $b < 0$ we see from (2.3) that the marginal utility is only positive if

$$E[x] < -\frac{1}{2b}. \tag{2.7}$$

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.[13]

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

$$E[U(x)] = E\left[x + bx^2\right] = \mu + bE\left[x^2\right] = \mu + b\left(\mu^2 + \sigma^2\right). \tag{2.8}$$

The indifference curves are obtained by totally differentiating both sides:

$$dE[U(x)] = (1 + 2b\mu)d\mu + 2\sigma d\sigma = 0. \tag{2.9}$$

The slope of the indifference curve in the (μ,σ)-plane is obtained by rearranging (2.9):[14]

$$\frac{d\mu}{d\sigma} = -\frac{2b\sigma}{1 + 2b\mu} = z\sigma > 0, \tag{2.10}$$

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 (2.8) for μ:[15]

$$\begin{aligned} E[U(x)] &= \mu + b\mu^2 + b\sigma^2 \\ \mu^2 + \frac{1}{b}\mu + \sigma^2 &= \frac{E[U(x)]}{b} \\ \left(\mu + \frac{1}{2b}\right)^2 + \sigma^2 &= \frac{1}{b}E[U(x)] + \frac{1}{4b^2}. \end{aligned} \tag{2.11}$$

[13] This restriction on the expected value is another argument often brought forward against the use of a quadratic utility function and hence the mean-variance criterion. Another argument is that the risk aversion increases with the expected outcome: $\frac{\partial z}{\partial E[x]} = \frac{4b^2}{(1+2bE[x])^2} > 0$, 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.

[14] 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.

[15] See Sharpe (1970, pp. 198 f.)

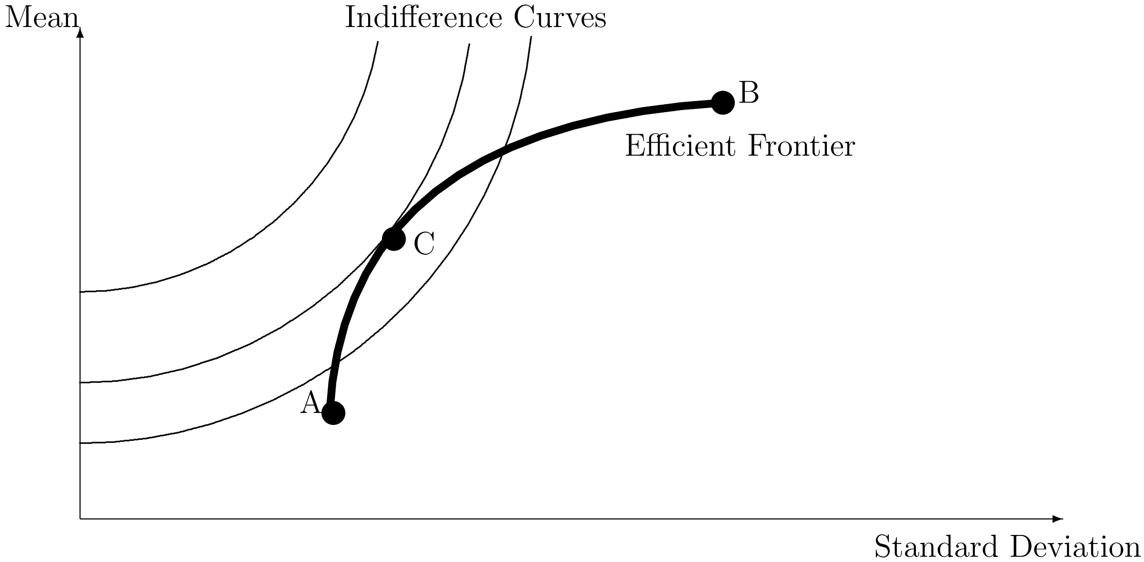

Fig. 2.3: Determination of the optimal alternative

Defining $r^* = -\frac{1}{2b}$ as the expected outcome that must not be exceeded for the marginal utility to be positive according to equation (2.7), we can rewrite the equation for the indifference curves as

$$(\mu - r^*) + \sigma^2 = -2r^* E\left[U(x)\right] + r^{*2} \equiv R^2, \tag{2.12}$$

which is the equation of a circle with center $\mu = r^*$, $\sigma = 0$ and radius R.[16] 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 2.3 shows the determination of the optimal alternative C.

We will now show that with a quadratic utility function the mean-variance criterion is optimal.[17] We assume two alternatives with $\mu_i = E\left[a_i\right] > E\left[a_j\right] = \mu_j$.

[16] 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 (2.1) we know that for a given σ the alternative with the higher expected outcome will be preferred. In practice this problem is overcome by using only the lower right sector of the circle.

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

Let further $\sigma_i^2 = Var[a_i]$ and $\sigma_j^2 = Var[a_j]$. If $a_i \succ a_j$ it has to be shown that

$$E[U(a_i)] > E[U(a_j)].$$

Substituting the utility functions gives

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

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

Dividing by $-2b > 0$ gives us

$$(\mu_i - \mu_j)\left[-\frac{1}{2b} - \frac{\mu_i + \mu_j}{2}\right] - \frac{\sigma_i^2 - \sigma_j^2}{2} > 0. \tag{2.13}$$

From (2.7) we know that $-\frac{1}{2b} > \mu_i$ and $-\frac{1}{2b} > \mu_j$, hence we find that

$$-\frac{1}{2b} > \frac{\mu_i + \mu_j}{2} \tag{2.14}$$

With the assumption that $\mu_i > \mu_j$ and as $b < 0$ the first term in (2.13) is positive. If now $\sigma_i^2 \leq \sigma_j^2$ as proposed by the mean-variance criterion equation (2.13) is fulfilled and we have shown that 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 (2.7). 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.[18]

2.2 The Markowitz frontier

The portfolio selection theory is based on several assumptions:[19]

- no transaction costs and taxes,
- assets are indefinitely divisible,

[18] 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 the literature as the Schneeweiss-Theorem.

[19] See Lintner (1965, p. 15).

- 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.

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 to choose between are the compositions of the portfolios, i.e. the weight each asset has.[20] Assume an investor has to choose between $N > 1$ assets, assigning a weight of x_i to each asset. The expected return of each asset is denoted μ_i and the variance of the returns by $\sigma_i^2 > 0$ for all $i = 1, \dots, N$.[21] The covariances between two assets i and j will be denoted σ_{ij}.

The weights of the assets 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:

$$\sum_{i=1}^{N} x_i = 1, \qquad x_i \geq 0, \; i = 1, \dots, N. \tag{2.15}$$

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.[22]

[20] 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 weights only. See Levy and Sarnat (1972, pp. 420 f.).

[21] 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 a transformation when using expected utility.

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

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

$$\mu_p = x_1\mu_1 + x_2\mu_2 = \mu_2 + x_1(\mu_1 - \mu_2), \tag{2.16}$$

$$\begin{aligned} \sigma_p^2 &= x_1^2\sigma_1^2 + x_2^2\sigma_2^2 + 2x_1x_2\sigma_{12} \\ &= \sigma_2^2 + x_1^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}) + 2x_1(\sigma_1\sigma_2\rho_{12} - \sigma_2^2), \end{aligned} \tag{2.17}$$

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 (2.17). 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 unless $\sigma_1 = \sigma_2$ 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 in the *minimum risk portfolio* (*MRP*):

$$x_1^{MRP} = \frac{\sigma_2^2 - \sigma_1\sigma_2\rho_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}. \tag{2.18}$$

The minimum variance can be obtained by inserting (2.18) into (2.17):

$$\begin{aligned} \sigma_{MRP}^2 &= \sigma_2^2 + \frac{(\sigma_2^2 - \sigma_1\sigma_2\rho_{12})^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}} - 2\frac{(\sigma_2^2 - \sigma_1\sigma_2\rho_{12})^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}} \\ &= \sigma_2^2 - \frac{(\sigma_2^2 - \sigma_1\sigma_2\rho_{12})^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}} \\ &= \frac{\sigma_1^2\sigma_2^2(1 - \rho_{12}^2)}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}. \end{aligned} \tag{2.19}$$

If the returns of the two assets are uncorrelated ($\rho_{12} = 0$), then (2.19) reduces to

$$\sigma_{MRP}^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \tag{2.20}$$

This variance is smaller than the variance of any of these two assets.[23] 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.

[23] 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 $\sigma_{MRP}^2 < \sigma_1^2$.

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

$$\sigma^2_{MRP} = 0 \tag{2.21}$$

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:

$$\sigma^2_{MRP} = \begin{cases} \sigma_1^2 & \text{if} \quad \sigma_1^2 \leq \sigma_2^2 \\ \sigma_2^2 & \text{if} \quad \sigma_1^2 > \sigma_2^2 \end{cases} . \tag{2.22}$$

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

$$\begin{aligned} \sigma_p^2 - \sigma^2_{MRP} &= \sigma_2^2 + x_1^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}) + 2x_1(\sigma_1\sigma_2\rho_{12} - \sigma_2^2) \\ &\quad - \frac{\sigma_1^2\sigma_2^2(1-\rho_{12}^2)}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}} \\ &= (x_1 - x_1^{MRP})^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}). \end{aligned} \tag{2.23}$$

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

$$\mu_p - \mu_{MRP} = (x_1 - x_1^{MRP})(\mu_1 - \mu_2). \tag{2.24}$$

Solving (2.24) for $x_1 - x_1^{MRP}$ and inserting into (2.23) we obtain after rearranging:

$$(\mu_p - \mu_{MPR})^2 = \frac{\sigma_p^2 - \sigma^2_{MRP}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}(\mu_1 - \mu_2)^2. \tag{2.25}$$

This equation represents a hyperbola with axes[24]

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

The efficient portfolios lie on the upper branch of this hyperbola, i.e. above the minimum risk portfolio.[25] Figure 2.4 shows the efficient portfolios for different correlations. It can easily be shown that in the case of perfect positive correlation

[24] See TOBIN (1966, p.30).

[25] The efficient frontier is also called *opportunity locus*.

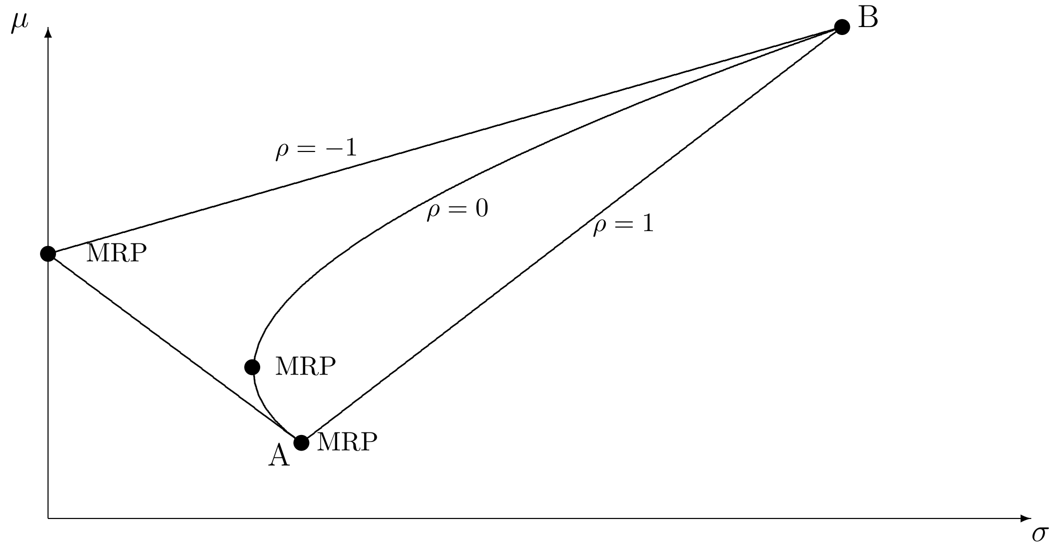

Fig. 2.4: Efficient portfolios with two assets

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 2.5 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 is 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 2.6 illustrates this situation. All achievable portfolio combinations are now located in the area bordered by the bold line connecting points A and C, where the bold line encircling the different hyperbolas is the new opportunity locus.

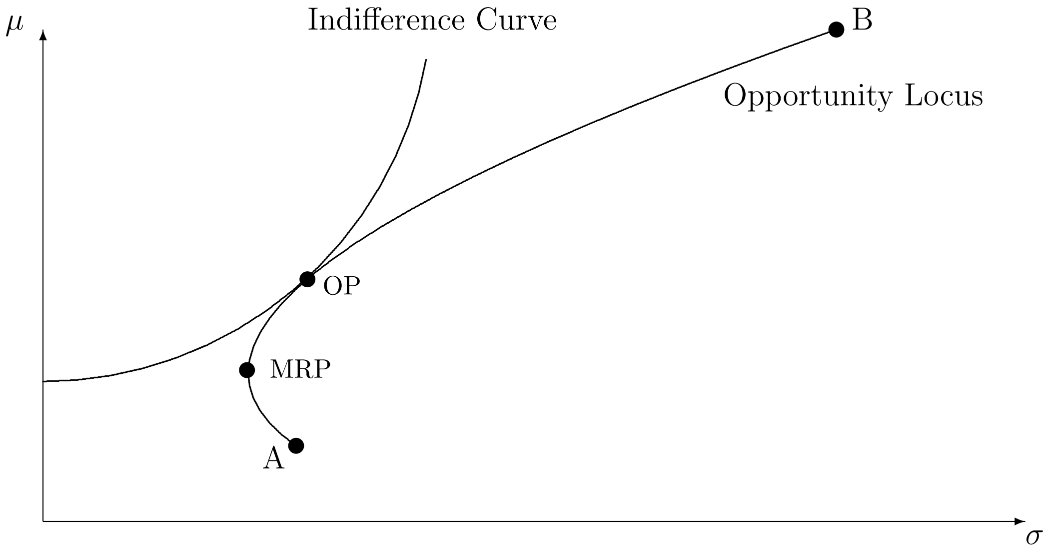

Fig. 2.5: Determination of the optimal portfolio with two assets

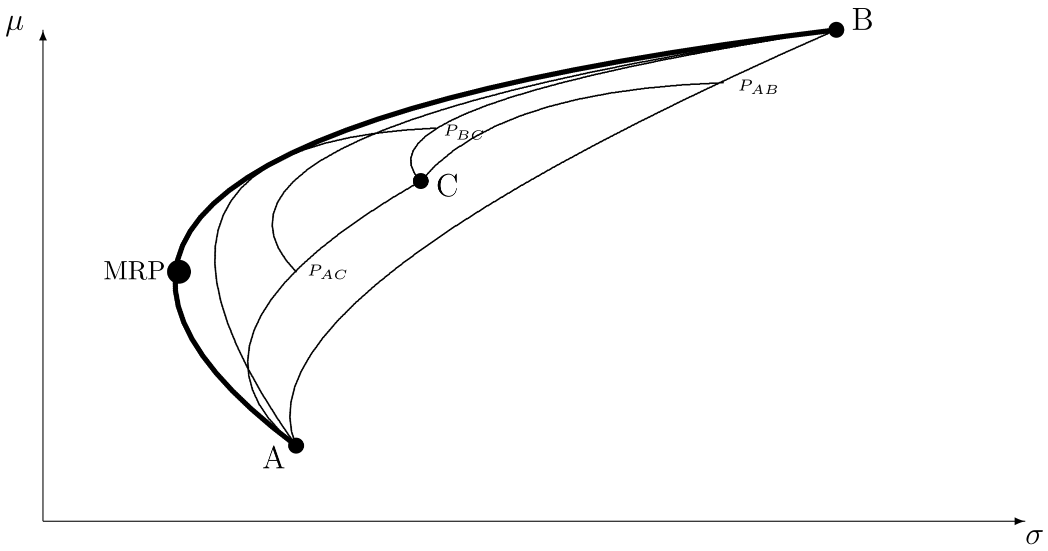

Fig. 2.6: Portfolio selection with three assets

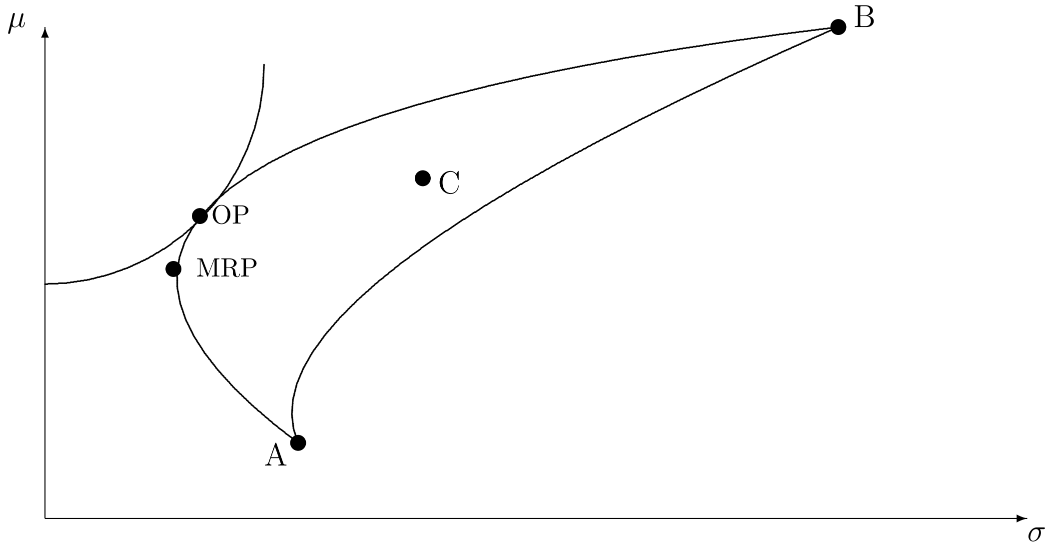

Fig. 2.7: The optimal portfolio with $N > 2$ assets

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 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 2.7. 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 case of two assets we get from (2.16) and (2.17):[26]

$$\mu_p = x_1\mu_1 + x_2 r = r + x_1(\mu_1 - r), \tag{2.26}$$

$$\sigma_p^2 = x_1^2\sigma_1^2. \tag{2.27}$$

Solving (2.27) for x_1 and inserting into (2.26) gives

$$\mu_p = r + \frac{\sigma_p}{\sigma_1}(\mu_1 - r) = r + \frac{\mu_1 - r}{\sigma_1}\sigma_p. \tag{2.28}$$

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 2.8 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.[27] 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 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

[26] See Tobin (1958).

[27] 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.

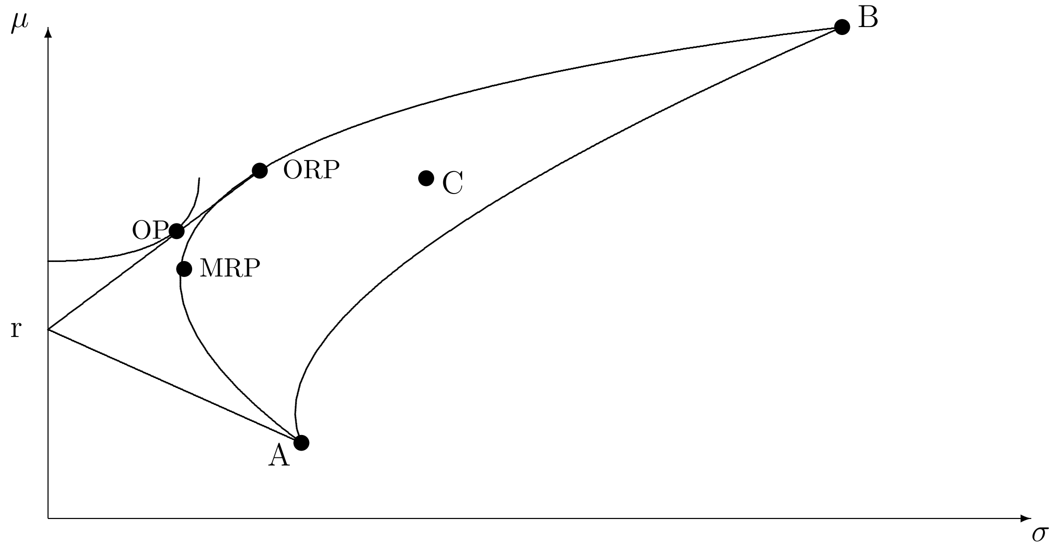

Fig. 2.8: The optimal portfolio with a riskless asset

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.*[28]

So far we have assumed that $x_i \geq 0$ for all $i = 1, \ldots, N$. 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 not 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 conditions (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

[28] 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 weights 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.

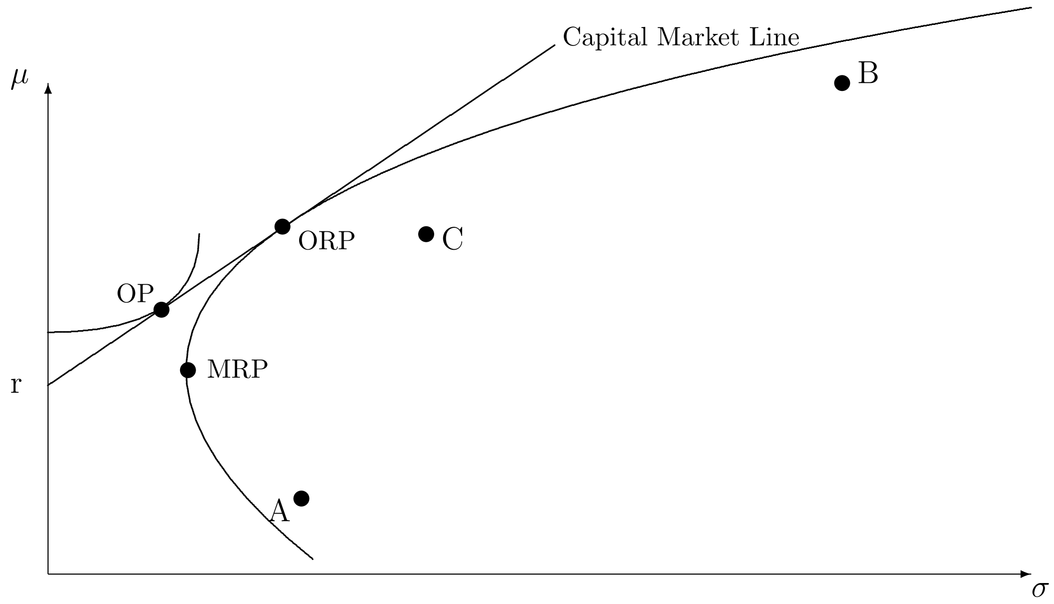

Fig. 2.9: Portfolio selection with short sales

2.9 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 larger 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.

To determine the efficient frontier and the optimal portfolio non-trivial numerical optimization routines have to be applied.[29] 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 appendix:

- the share to be invested into risky assets is determined by the optimal portfolio,
- the assets to invest in are those included in the optimal risky portfolio,
- the shares to invest in each selected asset are given by the weights of the optimal risky portfolio.

The portfolio theory has developed a method 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

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

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 in this framework.

Bibliography

Arrow, K. J. (1963): The Theory of Risk Aversion, chapter 9, 147–171.

Aschinger, G. (1990): General Remarks on Portfolio Theory and its Application. Working Paper Nr. 158, Insitute for Economic and Social Sciences, University of Fribourg (Switzerland).

Brachinger, H.-W. and Weber, M. (1997): Risk as a Primitive: A Survey of Measures of Perceived Risk. In: OR Spektrum, 19, 235–250.

Büschgen, H. (1991): Das kleine Börsen-Lexikon. 19th edition, Düsseldorf: Verlag Wirtschaft und Finanzen.

Cymbalista, F. (1998): Zur Unmöglichkeit rationaler Bewertung unter Unsicherheit - Eine monetär-keynesianische Kritik der Diskussion um die Markteffizienzthese. Studien zur Monetären Ökonomie. Marburg: Metropolis Verlag.

Dumas, B. and Allaz, B. (1996): Financial Securities: Market Equilibrium and Pricing Methods. London: Chapman & Hall.

Keynes, J. M. (1936): The General Theory of Employment, Interest, and Money. London.

Knight, F. H. (1957 (Orig. 1921)): Risk, Uncertainty and Profit. New York: Kelly & Millman, Inc.

Levy, H. and Sarnat, M. (1972): Investment and Portfolio Analysis. New York: John Wiley & Sons, Inc.

Lintner, J. (1965): Security Prices, Risk, and Maximal Gains from Diversification. In: Journal of Finance, 20, 587–615.

Markowitz, H. M. (1959): Portfolio Selection: Efficient Diversification of Investments. New Haven, CT: Yale University Press.

Pratt, J. W. (1964): Risk Aversion in the Small and in the Large. In: Econometrica, 32, 122–136.

Schumann, J. (1992): Grundzüge der mikroökonomischen Theorie. 6th edition, Berlin: Springer Verlag.

Sharpe, W. F. (1970): Portfolio Theory and Capital Markets. New York: McGraw-Hill Book Co.

Tobin, J. (1958): Liquidity Preference as a Behavior Towards Risk. In: Review of Economic Studies, 25, 65–86.

Tobin, J. (1966): The Theory of Portfolio Selection. In: F. H. Hahn and F. P. R. Brechling: The Theory of Interest Rates, New York, 3–51.

von Neumann, J. and Morgenstern, O. (1953): The Theory of Games and Economic Behavior. Princeton: Princeton University Press.