# Chapter 9

# Portfolio Theory

### Matrix Algebra

First we need a few things about matrices. (A very useful reference for mathematical results in the large class imprecisely defined as "well-known" is Berck & Sydsæter (1992), "Economists' Mathematical Handbook", Springer.)

- When $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{V} \in \mathbb{R}^{n \times n}$ then

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^\top \mathbf{V} \mathbf{x}) = (\mathbf{V} + \mathbf{V}^\top)\mathbf{x}$$

- A matrix $\mathbf{V} \in \mathbb{R}^{n \times n}$ is said to be *positive definite* if $\mathbf{z}^\top \mathbf{V} \mathbf{z} > 0$ for all $\mathbf{z} \neq \mathbf{0}$. If $\mathbf{V}$ is positive definite then $\mathbf{V}^{-1}$ exists and is also positive definite.

- Multiplying (appropriately) partitioned matrices is just like multiplying $2 \times 2$-matrices.

- When $X$ is an $n$-dimensional random variable with covariance matrix $\mathbf{\Sigma}$ then

$$\mathrm{Cov}(\mathbf{A}X + \mathbf{B}, \mathbf{C}X + \mathbf{D}) = \mathbf{A}\mathbf{\Sigma}\mathbf{C}^\top,$$

  where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are deterministic matrices such that the multiplications involved are well-defined.

### Basic Definitions & Justification of Mean-Variance Analysis

We will consider an agent who wants to invest in the financial markets. We look at a simple model with only two time-points, 0 and 1. The agent has an initial wealth of $W_0$ to invest. We are not interested in how the agent determined this amount, it's just there. There are $n$ financial assets to choose from and these have prices

$$S_{i,t} \text{ for } i = 1, \ldots, n \text{ and } t = 0, 1,$$

where $S_{i,1}$ is stochastic and not known until time 1. The rate of return on asset $i$ is defined as

$$r_i = \frac{S_{i,1} - S_{i,0}}{S_{i,0}},$$

and $r = (r_1, \ldots, r_n)^\top$ is the vector of rates of return. Note that $r$ is stochastic.

At time 0 the agent chooses a portfolio, that is he buys $a_i$ units of asset $i$ and since all in all $W_0$ is invested we have

$$W_0 = \sum_{i=1}^{n} a_i S_{i,0}.$$

(If $a_i < 0$ the agent is selling some of asset $i$; in most of our analysis short-selling will be allowed.)

Rather than working with the absolute number of assets held, it is more convenient to work with relative portfolio weights. This means that for the $i$th asset we measure the value of the investment in that asset relative to total investment and call this $w_i$, i.e.

$$w_i = \frac{a_i S_{i,0}}{\sum_{i=1}^{n} a_i S_{i,0}} = \frac{a_i S_{i,0}}{W_0}.$$

We put $\mathbf{w} = (w_1, \ldots w_n)^\top$, and have that $\mathbf{w}^\top \mathbf{1} = 1$. In fact, *any* vector satisfying this condition identifies an investment strategy. Hence in the following a portfolio is a vector whose coordinate sum to 1. Note that in this one period model a portfolio $\mathbf{w}$ is not a stochastic variable (in the sense of being unknown at time 0).

The terminal wealth is

$$\begin{aligned} W_1 &= \sum_{i=1}^{n} a_i S_{i,1} = \sum_{i=1}^{n} a_i (S_{i,1} - S_{i,0}) + \sum_{i=1}^{n} a_i S_{i,0} \\ &= W_0 \left( 1 + \sum_{i=1}^{n} \frac{S_{i,0} a_i}{W_0} \frac{S_{i,1} - S_{i,0}}{S_{i,0}} \right) \\ &= W_0 (1 + \mathbf{w}^\top r), \end{aligned} \tag{9.1}$$

so if we know the relative portfolio weights and the realized rates of return, we know terminal wealth. We also see that

$$E(W_1) = W_0 (1 + \mathbf{w}^\top E(r))$$

and

$$\mathrm{Var}(W_1) \;=\; W_0^2 \mathrm{Cov}(\mathbf{w}^\top r, \mathbf{w}^\top r) = W_0^2 \mathbf{w}^\top \mathrm{Var}(r) \mathbf{w}.$$

In this chapter we will look at how agents should choose $\mathbf{w}$. We will focus on how to choose $\mathbf{w}$ such that for a given expected rate of return, the variance on the rate of return is minimized. This is called mean-variance analysis. Intuitively, it sounds reasonable enough, but can it be justified?

An agent has a utility function, $u$, and let us for simplicity say that he derives utility from directly from terminal wealth. (So in fact we are saying that we can eat money.) We can expand $u$ in a Taylor series around the expected terminal wealth,

$$\begin{aligned} u(W_1) &= u(E(W_1)) + u'(E(W_1))(W_1 - E(W_1)) \\ &\quad + \frac{1}{2} u''(E(W_1))(W_1 - E(W_1))^2 + R_3, \end{aligned}$$

where the remainder term $R_3$ is

$$R_3 = \sum_{i=3}^{\infty} \frac{1}{i!} u^{(i)}(E(W_1))(W_1 - E(W_1))^i,$$

"and hopefully small". With appropriate (weak) regularity condition this means that expected terminal wealth can be written as

$$E(u(W_1)) = u(E(W_1)) + \frac{1}{2} u''(E(W_1)) \mathrm{Var}(W_1) + E(R_3),$$

where the remainder term involves higher order central moments. As usual we consider agents with increasing, concave (i.e. $u'' < 0$) utility functions who maximize expected wealth. This then shows that to a second order approximation there is a preference for expected wealth (and thus, by (9.1), to expected rate of return), and an aversion towards variance of wealth (and thus to variance of rates of return).

But we also see that mean-variance analysis cannot be a completely general model of portfolio choice. A sensible question to ask is: What restrictions can we impose (on $u$ and/or on $r$) to ensure that mean-variance analysis is fully consistent with maximization of expected utility?

An obvious way to do this is to assume that utility is quadratic. Then the remainder term is identically 0. But quadratic utility does not go too well with the assumption that utility is increasing and concave. If $u$ is concave (which it has to be for mean-variance analysis to hold ; otherwise our interest would be in maximizing variance) there will be a point of satiation beyond which utility decreases. Despite this, quadratic utility is often used with a "happy-go-lucky" assumption that when maximizing, we do not end up in an area where it is decreasing.

We can also justify mean-variance analysis by putting distributional restrictions on rates of return. If rates of return on individual assets are normally distributed then the rate of return on a portfolio is also normal, and the higher order moments in the remainder can be expressed in terms of the variance. In general we are still not sure of the signs and magnitudes of the higher order derivatives of $u$, but for large classes of reasonable utility functions, mean-variance analysis can be formally justified.

## 9.1 The Mathematics of the Efficient Frontier

### 9.1.1 The case with no riskfree asset

First we consider a market with no riskfree asset and $n$ risky assets. Later we will include a riskfree asset, and it will become apparent that we have done things in the right order.

The risky assets have a vector of rates of return of $r$, and we assume that

$$\begin{aligned} E(r) &= \boldsymbol{\mu}, && (9.2)\\ \mathrm{Var}(r) &= \boldsymbol{\Sigma}, && (9.3) \end{aligned}$$

where $\boldsymbol{\Sigma}$ is positive definite (hence invertible) and not all coordinates of $\boldsymbol{\mu}$ are equal. As a covariance matrix $\boldsymbol{\Sigma}$ is always positive semidefinite, the definiteness means that there does not exist an asset whose rate of return can be written as an affine function of the other $n-1$ assets' rates of return. Note that the existence of a riskfree asset would violate this.

Consider the following problem:

$$\min_{\mathbf{w}} \frac{1}{2}\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} := \sigma_P^2 \quad \text{subject to} \quad \begin{aligned} \mathbf{w}^\top \boldsymbol{\mu} &= r_P \\ \mathbf{w}^\top \mathbf{1} &= 1 \end{aligned}$$

First note that our assumptions on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ ensure that a unique finite solution exits for any value of $r_P$. Second note that the problem can be interpreted as choosing portfolio weights (the second constraint ensures that $\mathbf{w}$ is a vector of portfolio weights) such that the variance on the return on the portfolio ($\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}$; the "1/2" is just there for convenience) is minimized given that we want a specific expected rate of return ($r_P$; "$P$ is for portfolio").

To solve the problem we set up the Lagrange-function with multipliers

$$\mathcal{L}(\mathbf{w}, \lambda_1, \lambda_2) = \frac{1}{2}\mathbf{w}^\top \mathbf{\Sigma} \mathbf{w} - \lambda_1(\mathbf{w}^\top \boldsymbol{\mu} - r_P) - \lambda_2(\mathbf{w}^\top \mathbf{1} - 1).$$

The first-order conditions for optimality are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \mathbf{\Sigma}\mathbf{w} - \lambda_1 \boldsymbol{\mu} - \lambda_2 \mathbf{1} &= 0, && (9.4) \\ \mathbf{w}^\top \boldsymbol{\mu} - r_P &= 0, && (9.5) \\ \mathbf{w}^\top \mathbf{1} - 1 &= 0. && (9.6) \end{aligned}$$

Usually we might say "and these are linear equations that can easily be solved", but working on them algebraically leads to a much deeper understanding and intuition about the model. Note that invertibility gives that we can write (9.4) as (check for yourself)

$$\mathbf{w} = \mathbf{\Sigma}^{-1}[\boldsymbol{\mu}\ \ \mathbf{1}] \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}, \tag{9.7}$$

and (9.5)-(9.6) as

$$[\boldsymbol{\mu}\ \ \mathbf{1}]^\top \mathbf{w} = \begin{bmatrix} r_P \\ 1 \end{bmatrix}. \tag{9.8}$$

Multiplying both sides of (9.7) by $[\boldsymbol{\mu}\ \ \mathbf{1}]^\top$ and using (9.8) gives

$$\begin{bmatrix} r_P \\ 1 \end{bmatrix} = [\boldsymbol{\mu}\ \mathbf{1}]^\top \mathbf{w} = \underbrace{[\boldsymbol{\mu}\ \mathbf{1}]^\top \mathbf{\Sigma}^{-1} [\boldsymbol{\mu}\ \mathbf{1}]}_{:=\mathbf{A}} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}. \tag{9.9}$$

By using the multiplication rules for partitioned matrices we see that

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{\mu}^\top \mathbf{\Sigma}^{-1} \boldsymbol{\mu} & \boldsymbol{\mu}^\top \mathbf{\Sigma}^{-1} \mathbf{1} \\ \boldsymbol{\mu}^\top \mathbf{\Sigma}^{-1} \mathbf{1} & \mathbf{1}^\top \mathbf{\Sigma}^{-1} \mathbf{1} \end{bmatrix} := \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

We now show that $\mathbf{A}$ is positive definite, in particular it is invertible. To this end let $\mathbf{z}^\top = (z_1, z_2) \neq \mathbf{0}$ be an arbitrary non-zero vector in $\mathbb{R}^2$. Then

$$\mathbf{y} = [\boldsymbol{\mu}\ \mathbf{1}] \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = [z_1 \boldsymbol{\mu}\ \ \ z_2 \mathbf{1}] \neq \mathbf{0},$$

because the coordinates of $\boldsymbol{\mu}$ are not all equal. From the definition of $\mathbf{A}$ we get

$$\forall \mathbf{z} \neq \mathbf{0} \ \ : \ \ \mathbf{z}^\top \mathbf{A} \mathbf{z} = \mathbf{y}^\top \mathbf{\Sigma}^{-1} \mathbf{y} > 0,$$

because $\mathbf{\Sigma}^{-1}$ is positive definite (because $\mathbf{\Sigma}$ is). In other words, $\mathbf{A}$ is positive definite. Hence we can solve (9.9) for the $\lambda$'s,

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} r_P \\ 1 \end{bmatrix},$$

and insert this into (9.7) in order to determine the optimal portfolio weights

$$\widehat{\mathbf{w}} = \mathbf{\Sigma}^{-1}[\boldsymbol{\mu}\ \mathbf{1}]\mathbf{A}^{-1} \begin{bmatrix} r_P \\ 1 \end{bmatrix}. \tag{9.10}$$

The portfolio $\widehat{\mathbf{w}}$ is called the minimum variance portfolio for a given mean $r_P$ (So we can't be bothered to say the correct full phrase: "minimum variance on rate of return for a given mean rate on return $r_P$ ".) Twice the optimal value (i.e. the minimal portfolio return variance) is

$$\begin{aligned} \widehat{\sigma}_P^2 &= \widehat{\mathbf{w}}^\top \mathbf{\Sigma} \widehat{\mathbf{w}} \\ &= [r_P\ \ 1]\mathbf{A}^{-1}[\boldsymbol{\mu}\ \ \mathbf{1}]^\top \mathbf{\Sigma}^{-1}\mathbf{\Sigma}\mathbf{\Sigma}^{-1}[\boldsymbol{\mu}\ \ \mathbf{1}]\mathbf{A}^{-1}[r_P\ \ 1]^\top \\ &= [r_P\ \ 1]\mathbf{A}^{-1} \underbrace{\left([\boldsymbol{\mu}\ \ \mathbf{1}]^\top \mathbf{\Sigma}^{-1}[\boldsymbol{\mu}\ \ \mathbf{1}]\right)}_{=\mathbf{A}\ \text{ by def.}} \mathbf{A}^{-1}[r_P\ \ 1]^\top \\ &= [r_P\ \ 1]\mathbf{A}^{-1} \begin{bmatrix} r_P \\ 1 \end{bmatrix}, \end{aligned}$$

where symmetry (of $\mathbf{\Sigma}$ and $\mathbf{A}$ and their inverses) was used to obtain the second line. But note that

$$\mathbf{A}^{-1} = \frac{1}{ac - b^2} \begin{bmatrix} c & -b \\ -b & a \end{bmatrix},$$

which gives us

$$\widehat{\sigma}_P^2 = \frac{a - 2br_P + cr_P^2}{ac - b^2}. \tag{9.11}$$

In (9.11) the relation between the variance of the minimum variance portfolio for a given $r_p$, $\widehat{\sigma}_P^2$, is expressed as a parabola and is called the *variance portfolio frontier* or *locus*. In mean-standard deviation-space the relation is expressed as a hyperbola. Figure 9.1 illustrates what things look like in mean-variance-space. (When using graphical arguments you should be quite careful to use "the right space"; for instance lines that are straight in one space, are not straight in the other.) The upper half of the curve in Figure 9.1 (the solid line) identifies the set of portfolios that have the highest mean return for a given variance; these are called mean-variance *efficient portfolios*. The portfolios on the bottom half (the dotted part) are called *inefficient*

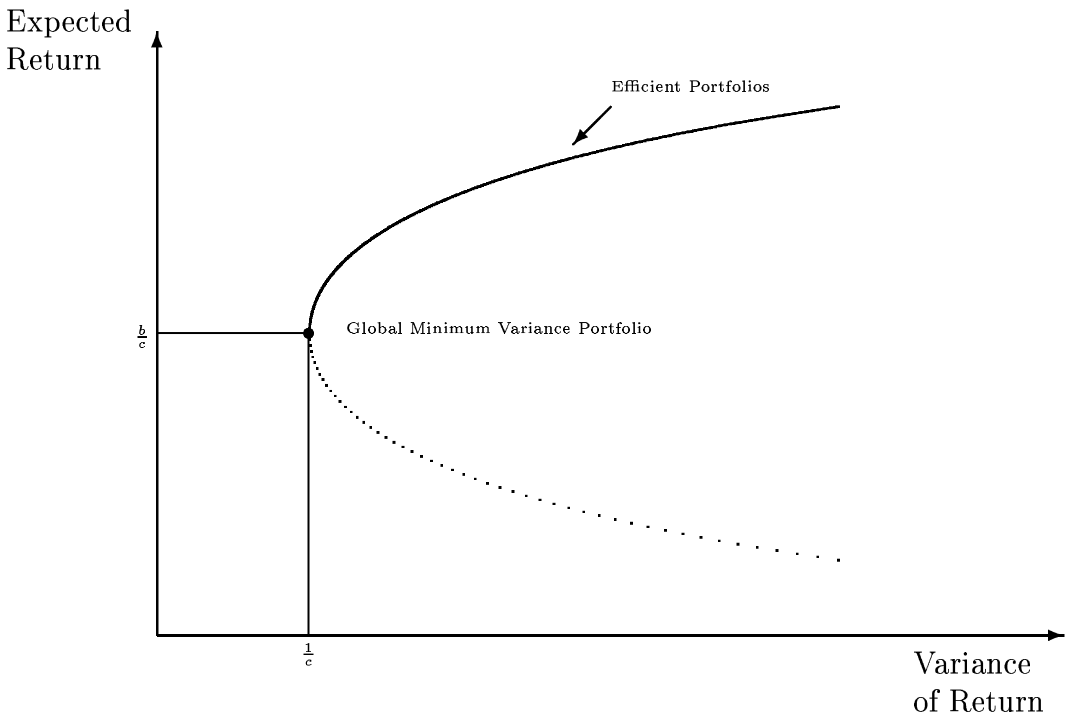


Figure 9.1: The minimum variance portfolio frontier.

*portfolios*. Figure 9.1 also shows the *global minimum variance portfolio*, the portfolio with the smallest possible variance for any given mean return. Its mean, $r_G$, is found by minimizing (9.11) with respect to $r_P$, and is $r_{gmv} = \frac{b}{c}$. By substituting this in the general $\widehat{\sigma}^2$-expression we obtain

$$\widehat{\sigma}^2_{gmv} = \frac{a - 2br_{gmv} + cr^2_{gmv}}{ac - b^2} = \frac{a - 2b(b/c) + c(b/c)^2}{ac - b^2} = \frac{1}{c},$$

while the general formula for portfolio weights gives us

$$\widehat{\mathbf{w}}_{gmv} = \frac{1}{c}\mathbf{\Sigma}^{-1}\mathbf{1}.$$

**Example 11 (A Recurrent Numerical Example)** Consider the case with 3 assets (referred to as $A$, $B$, and $C$) and

$$\boldsymbol{\mu} = \begin{bmatrix} 0.1 \\ 0.12 \\ 0.15 \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} 0.25 & 0.10 & -0.10 \\ 0.10 & 0.36 & -0.30 \\ -0.10 & -0.30 & 0.49 \end{bmatrix}.$$

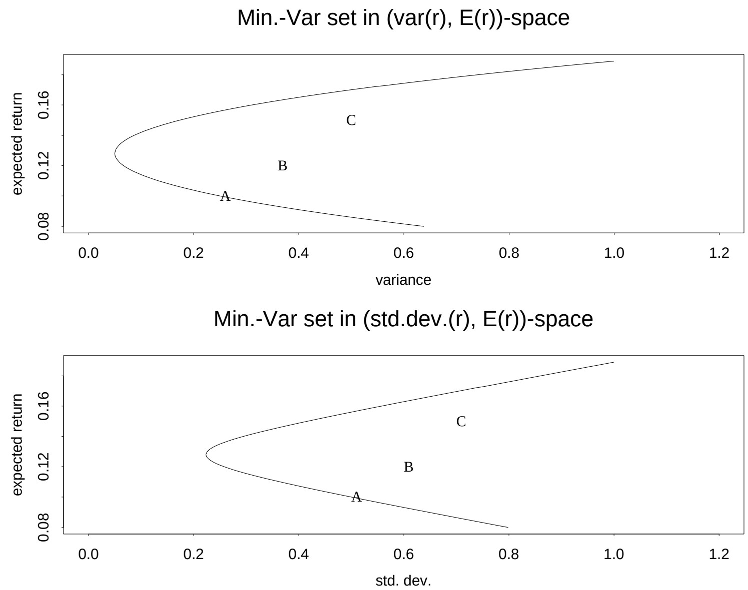


Figure 9.2: The minimum variance frontiers and individual assets for Example 11

The all-important **A**-matrix is then

$$\mathbf{A} = \begin{bmatrix} 0.33236 & 2.56596 \\ 2.565960 & 20.04712 \end{bmatrix},$$

which means that the locus of mean-variance portfolios is given by

$$\widehat{\sigma}_P^2 = 4.22918 - 65.3031 r_P + 255.097 r_P^2.$$

The locus is illustrated in Figure 9.2 in both in (variance, expected return)-space and (standard deviation, expected return)-space.

An important property of the set of minimum variance portfolios is is so-called two-fund separation. This means that the minimum variance portfolio frontier can be generated by any two distinct frontier portfolios.

**Proposition 30** *Let $\mathbf{x}_a$ and $\mathbf{x}_b$ be two minimum variance portfolios with mean returns $r_a$ and $r_b$, $r_a \neq r_b$. Then every minimum variance portfolio, $\mathbf{x}_c$ is a linear combination of $\mathbf{x}_a$ and $\mathbf{x}_b$. Conversely, every portfolio that is a linear combination of $\mathbf{x}_a$ and $\mathbf{x}_b$ (i.e. can be written as $\alpha\mathbf{x}_a + (1-\alpha)\mathbf{x}_b$) is a minimum variance portfolio. In particular, if $\mathbf{x}_a$ and $\mathbf{x}_b$ are efficient portfolios, then $\alpha\mathbf{x}_a + (1-\alpha)\mathbf{x}_b$ is an efficient portfolio for $\alpha \in [0; 1]$.*

**Proof**. To prove the first part let $r_c$ denote the mean return on a given minimum variance portfolio $\mathbf{x}_c$. Now choose $\alpha$ such that $r_c = \alpha r_a + (1-\alpha) r_b$, that is $\alpha = (r_c - r_b)/(r_a - r_b)$ (which is well-defined because $r_a \neq r_b$). But since $\mathbf{x}_c$ is a minimum variance portfolio we know that (9.10) holds, so

$$\begin{aligned} \mathbf{x}_c &= \boldsymbol{\Sigma}^{-1}[\boldsymbol{\mu} \;\; \mathbf{1}]\mathbf{A}^{-1} \begin{bmatrix} r_c \\ 1 \end{bmatrix} \\ &= \boldsymbol{\Sigma}^{-1}[\boldsymbol{\mu} \;\; \mathbf{1}]\mathbf{A}^{-1} \begin{bmatrix} \alpha r_a + (1-\alpha) r_b \\ \alpha + (1-\alpha) \end{bmatrix} \\ &= \alpha \mathbf{x}_a + (1-\alpha)\mathbf{x}_b, \end{aligned}$$

where the third line is obtained because $\mathbf{x}_a$ and $\mathbf{x}_b$ also fulfill (9.10). This proves the first statement. The second statement is proved by "reading from right to left" in the above equations. This shows that $\mathbf{x}_c = \alpha \mathbf{x}_a + (1-\alpha)\mathbf{x}_b$ is the minimum variance portfolio with expected return $\alpha r_a + (1-\alpha) r_b$. From this, the validity of the third statement is clear. ■

Another important notion is *orthogonality* of portfolios. We say that two portfolios $\mathbf{x}_P$ and $\mathbf{x}_{zP}$ ("$z$ is for zero") are orthogonal if the covariance of their rates of return is 0, i.e.

$$\mathbf{x}_{zP}^\top \boldsymbol{\Sigma} \mathbf{x}_P = 0. \tag{9.12}$$

Often $\mathbf{x}_{zP}$ is called $\mathbf{x}_P$'s 0-$\beta$ portfolio (we'll see why later).

**Proposition 31** *For every minimum variance portfolio, except the global minimum variance portfolio, there exists a unique orthogonal minimum variance portfolio. Furthermore, if the first portfolio has mean rate of return $r_P$, its orthogonal one has mean*

$$r_{zP} = \frac{a - b r_P}{b - c r_P}.$$

**Proof**. First note that $r_{zP}$ is well-defined for any portfolio except the global minimum variance portfolio. By (9.10) we know how to find the minimum variance portfolios with means $r_P$ and $r_{zP} = (a - b r_P)/(b - c r_P)$. This leads to

$$\begin{aligned} \mathbf{x}_{zP}^\top \boldsymbol{\Sigma} \mathbf{x}_P &= [r_{zP} \;\; 1]\mathbf{A}^{-1}[\boldsymbol{\mu} \;\; \mathbf{1}]^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}[\boldsymbol{\mu} \;\; \mathbf{1}]\mathbf{A}^{-1}[r_P \;\; 1]^\top \\ &= [r_{zP} \;\; 1]\mathbf{A}^{-1} \underbrace{\left([\boldsymbol{\mu} \;\; \mathbf{1}]^\top \boldsymbol{\Sigma}^{-1} [\boldsymbol{\mu} \;\; \mathbf{1}]\right)}_{=\mathbf{A} \text{ by def.}} \mathbf{A}^{-1}[r_P \;\; 1]^\top \end{aligned}$$

$$
\begin{aligned}
&= [r_{zP} \;\; 1]\mathbf{A}^{-1}\begin{bmatrix} r_P \\ 1 \end{bmatrix} && (9.13)\\
&= \begin{bmatrix} \frac{a-br_P}{b-cr_P} & 1 \end{bmatrix} \frac{1}{ac-b^2}\begin{bmatrix} c & -b \\ -b & a \end{bmatrix}\begin{bmatrix} r_P \\ 1 \end{bmatrix}\\
&= \frac{1}{ac-b^2}\begin{bmatrix} \frac{a-br_P}{b-cr_P} & 1 \end{bmatrix}\begin{bmatrix} cr_P - b \\ a - br_P \end{bmatrix}\\
&= 0,
\end{aligned}
$$

which was the desired result. ■

**Proposition 32** *Let* $\mathbf{x}_{mv}$ *(*$\neq \mathbf{x}_{gmv}$*, the global minimum variance portfolio) be a portfolio on the mean-variance frontier with rate of return* $r_{mv}$*, expected rate of return* $\mu_{mv}$ *and variance* $\sigma^2_{mv}$*. Let* $\mathbf{x}_{zmv}$ *be the corresponding orthogonal portfolio,* $\mathbf{x}_P$ *be an arbitrary portfolio, and use similar notation for rates of return on these portfolios. Then the following holds:*

$$\mu_P - \mu_{zmv} = \beta_{P,mv}(\mu_{mv} - \mu_{zmv}),$$

*where*

$$\beta_{P,mv} = \frac{Cov(r_P, r_{mv})}{\sigma^2_{mv}}.$$

**Proof**. Consider first the covariance between return on asset $i$ and $\mathbf{x}_{mv}$. By using (9.10) we get

$$
\begin{aligned}
\mathrm{Cov}(r_i, r_{mv}) &= \mathbf{e}_i^\top \boldsymbol{\Sigma}\mathbf{x}_{mv}\\
&= \mathbf{e}_i^\top[\boldsymbol{\mu} \;\; \mathbf{1}]\mathbf{A}^{-1}\begin{bmatrix} \mu_{mv} \\ 1 \end{bmatrix}\\
&= [\mu_i \;\; 1]\mathbf{A}^{-1}\begin{bmatrix} \mu_{mv} \\ 1 \end{bmatrix}.
\end{aligned}
$$

From calculations in the proof of Proposition 31 we know that the covariance between $\mathbf{x}_{mv}$ and $\mathbf{x}_{zvp}$ is given by (9.13). We also know that it is 0. Subtracting this 0 from the above equation gives

$$
\begin{aligned}
\mathrm{Cov}(r_i, r_{mv}) &= [\mu_i - \mu_{zmv} \;\; 0]\mathbf{A}^{-1}\begin{bmatrix} \mu_{mv} \\ 1 \end{bmatrix}\\
&= (\mu_i - \mu_{zmv})\underbrace{\frac{c\mu_{mv} - b}{ac-b^2}}_{:=\gamma}, && (9.14)
\end{aligned}
$$

where we have used the formula for $\mathbf{A}^{-1}$. Since this holds for all individual assets and covariance is bilinear, it also holds for portfolios. In particular for $\mathbf{x}_{mv}$,

$$\sigma_{mv}^2 = \gamma(\mu_{mv} - \mu_{zmv}),$$

so $\gamma = \sigma_{mv}^2/(\mu_{mv} - \mu_{zmv})$. By substituting this into (9.14) we get the desired result for individual assets. But then linearity ensures that it holds for all portfolios. ■

Proposition 32 says that the expected excess return on any portfolio (over the expected return on a certain portfolio) *is a linear function* of the expected excess return on a minimum variance portfolio. It also says that the expected excess return is proportional to covariance.

### 9.1.2 The case with a riskfree asset

We now consider a portfolio selection problem with $n+1$ assets. These are indexed by $0, 1, \ldots, n$, and 0 corresponds to the riskfree asset with (deterministic) rate of return $r_0$. For the risky assets we let ${r_i}^e$ denote the *excess* rate of return over the riskfree asset, i.e. the actual rate of return less $r_0$. We let $\boldsymbol{\mu}^e$ denote the mean excess rate of return, and $\boldsymbol{\Sigma}$ the variance (which is of course unaffected). A portfolio is now a $n+1$-dimensional vector whose coordinate sum to unity. But in the calculations we let $\mathbf{w}$ denote the vector of weights $w_1, \ldots, w_n$ corresponding to the risky assets and write $w_0 = 1 - \mathbf{w}^\top \mathbf{1}$.

With these conventions the mean excess rate of return on a portfolio $P$ is

$$r_P^e = \mathbf{w}^\top \boldsymbol{\mu}^e$$

and the variance is

$$\sigma_P^2 = \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}.$$

Therefore the mean-variance portfolio selection problem with a riskless asset can be stated as

$$\min_{\mathbf{w}} \frac{1}{2} \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^\top \boldsymbol{\mu}^e = r_P^e.$$

Note that $\mathbf{w}^\top \mathbf{1} = 1$ is not a constraint; some wealth may be held in the riskless asset.

As in the previous section we can set up the Lagrange-function, differentiate it, at solve to first order conditions. This gives the optimal weights

$$\widehat{\mathbf{w}} = \frac{r_P^e}{(\boldsymbol{\mu}^e)^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^e} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^e, \tag{9.15}$$

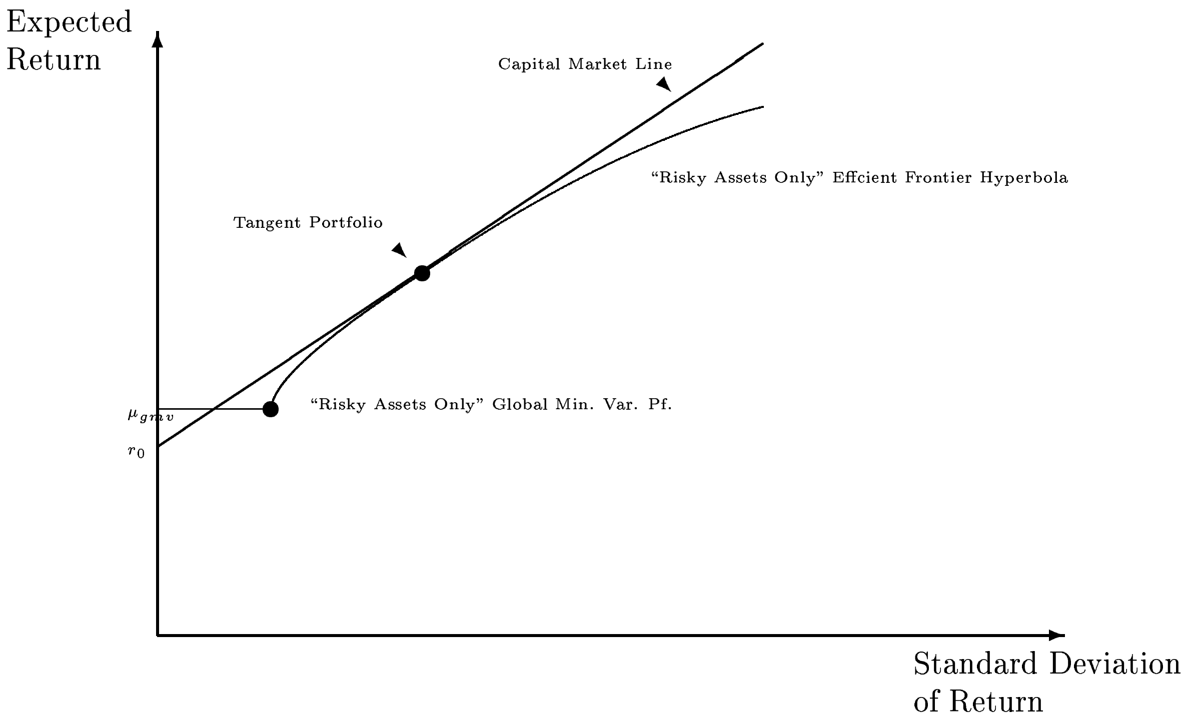


Figure 9.3: The capital market line.

and the following expression for the variance of the minimum variance portfolio with mean excess return $r_P$:

$$\widehat{\sigma}_P^2 = \frac{(r_P^e)^2}{(\boldsymbol{\mu}^e)^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^e}. \tag{9.16}$$

So we have determined the efficient frontier. For required returns above the riskfree rate, the efficient frontier in standard deviation-mean space is a straight line passing through $(0, r_0)$ with a slope of $\sqrt{(\boldsymbol{\mu}^e)^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^e}$. This line is called the capital market line.

The tangent portfolio, $\mathbf{x}$, is the minimum variance portfolio with all wealth invested in the risky assets, i.e. $\mathbf{x}_{tan}^\top \mathbf{1} = 1$. The mean excess return on the tangent portfolio is

$$r_{tan}^e = \frac{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}},$$

which may be positive or negative. It is economically plausible to assert that the riskless return is lower than the mean return of the global minimum variance portfolio of the risky assets. In this case the situation is as illustrated in Figure 9.3, and that explains why we use the term "tangency". When $r^e_{tan} > 0$, the tangent portfolio is on the capital market line. But the tangent portfolio must also be on the "risky assets only" efficient frontier. So the straight line (the CML) and the hyperbola intersect at a point corresponding to the tangency portfolio. But clearly the CML must be above the efficient frontier hyperbola (we are minimizing variance with an extra asset). So the CML is a tangent to the hyperbola.

For any portfolio, $P$ we define the *Sharpe-ratio* as excess return relative to standard deviation,

$$\text{Sharpe-ratio}_P = \frac{\mu_P - r_0}{\sigma_P}.$$

In the case where $r^e_{tan} > 0$, we see note from Figure 9.3 that the tangency portfolio is the "risky assets only"-portfolio with the highest Sharpe-ratio since the slope of the CML is the Sharpe-ratio of tangency portfolio. (Generally/"strictly algebraically" we should say that $\mathbf{x}_{tan}$ has maximal squared Sharpe-ratio.)

Note that a portfolio with full investment in the riskfree asset is orthogonal to any other portfolio; this means that we can prove the following result in exactly the manner as Proposition 32.

**Proposition 33** *Let $\mathbf{x}_{mv}$ be a portfolio on the mean-variance frontier with rate of return $r_{mv}$, expected rate of return $\mu_{mv}$ and variance $\sigma^2_{mv}$. Let $\mathbf{x}_P$ be an arbitrary portfolio, and use similar notation for rates of return on these portfolios. Then the following holds:*

$$\mu_P - r_0 = \beta_{P,mv}(\mu_{mv} - r_0),$$

*where*

$$\beta_{P,mv} = \frac{Cov(r_P, r_{mv})}{\sigma^2_{mv}}.$$

## 9.2 The Capital Asset Pricing Model (CAPM)

With the machinery of portfolio optimization in place, we are ready to formulate one of the key results of modern finance theory, the CAPM-relation. Despite the clearly unrealistic assumptions on which the result is built it still provides invaluable intuition on what factors determine the price of assets in equilibrium. Note that until now, we have mainly been concerned with pricing (derivative) securities when taking prices of some basic securities as given.

Here we try to get more insight into what determines prices of securities to begin with.

We consider an economy with $n$ risky assets and one riskless asset. Here, we let $r_i$ denote the rate of return on the $i$'th risky asset and we let $r_0$ denote the riskless rate of return. We assume that $r_0$ is strictly smaller than the return of the global minimum variance portfolio.

Just as in the case of only risky assets one can show that with a riskless asset the expected return on any asset or portfolio can be expressed as a function of its beta with respect to an efficient portfolio. In particular, since the tangency portfolio is efficient we have

$$Er_i - r_0 = \beta_{i,tan}(Er_{tan} - r_0) \tag{9.17}$$

where

$$\beta_{i,tan} = \frac{Cov(r_i, r_{tan})}{\sigma^2_{tan}} \tag{9.18}$$

The critical component in deriving the CAPM is the identification of the tangency portfolio as the *market portfolio.* The market portfolio is defined as follows: Assume that the initial supply of risky asset $j$ at time 0 has a value of $P_0^j$. (So $P_0^j$ is the number of shares outstanding times the price per share.) The market portfolio of risky assets then has portfolio weights given as

$$w_j^m = \frac{P_j^0}{\sum_{i=1}^n P_i^0} \tag{9.19}$$

Note that it is quite reasonable to think of a portfolio with these weights as reflecting "the average of the stock market".

Now if all (say $K$) agents are mean-variance optimizers (given wealths of $W_i(0)$ to invest), we know that since there is a riskless asset they will hold a combination of the tangency portfolio and the riskless asset since two fund separation applies. Hence all agents must hold the same mix of risky assets as that of the tangency portfolio. This in turn means that in equilibrium where market clearing requires all the risky assets to be held, the market portfolio (which is a convex combination of the individual agents' portfolios) has the same mixture of assets as the tangency portfolio. Or in symbols: Let $\phi_i$ denote the fraction of his wealth that agent $i$ has invested in the tangency portfolio. By summing over all agents we get

$$\begin{aligned}
\text{Total value of asset } j &= \sum_{i=1}^K \phi_i W_i(0)\mathbf{x}_{tan}(j) \\
&= \mathbf{x}_{tan}(j) \times \text{Total value of all risky assets,}
\end{aligned}$$

where we have used that market clearing condition that all risky assets must be held by the agents. (This is a very weak consequence of equilibrium; some would just call it an accounting identity. The main *economic assumption* is that agents are mean-variance optimizers so that two fund separation applies.) Hence we may as well write the market portfolio in equation (9.17). This is the CAPM:

$$Er_i - r_0 = \beta_{i,m}(Er_m - r_0) \tag{9.20}$$

where $\beta_{i,m}$ is defined using the market portfolio instead of the tangency portfolio. Note that the type of risk for which agents receive excess returns are those that are correlated with the market. The intuition is as follows: If an asset pays off a lot when the economy is wealthy (i.e. when the return of the market is high) that asset contributes wealth in states where the marginal utility of receiving extra wealth is small. Hence agents are not willing to pay very much for such an asset at time 0. Therefore, the asset has a high return. The opposite situation is also natural at least if one ever considered buying insurance: An asset which moves opposite the market has a high pay off in states where marginal utility of receiving extra wealth is high. Agents are willing to pay a lot for that at time 0 and therefore the asset has a low return. Indeed it is probably the case that agents are willing to accept a return on an insurance contract which is below zero. This gives the right intuition but the analogy with insurance is actually not completely accurate in that the risk one is trying to avoid by buying an insurance contract is not linked to market wide fluctuations.

Note that one could still view the result as a sort of relative pricing result in that we are pricing everything in relation to the given market portfolio. To make it more clear that there is an equilibrium type argument underlying it all, let us see how characteristics of agents help in determining the risk premium on the market portfolio. Consider the problem of agent $i$ in the one period model. We assume that returns are multivariate normal and that the utility function is twice differentiable and concave[1]:

$$\begin{aligned}&\max_{\mathbf{w}} E(u_i(W_1^i))\\ &\text{s.t.} W_1^i = W_0(\mathbf{w}^\top \mathbf{r} + (1 - \mathbf{w}^\top \mathbf{1})r_0)\end{aligned}$$

When forming the Lagrangian of this problem, we see that the first order condition for optimality is that for each asset $j$ and each agent $i$ we have

$$E\left(u_i'(W_1^i)(r_j - r_0)\right) = 0 \tag{9.21}$$

[1] This derivation follows Huang and Litzenberger: *Foundations for Financial Economics*

Remembering that $Cov(X,Y) = EXY - EXEY$ we rewrite this as

$$E\left(u_i'(W_1^i)\right) E(r_j - r_0) = -Cov(u_i'(W_1^i), r_j)$$

A nice lemma known as Stein's lemma says that for bivariate normal distribution $(X,Y)$ we have

$$Cov(g(X), Y) = Eg'(X)Cov(X,Y)$$

and using this we have the following first order condition:

$$E\left(u_i'(W_1^i)\right) E(r_j - r_0) = -Eu_i''(W_1^i)Cov(W_1^i, r_j)$$

i.e.

$$\frac{-E\left(u_i'(W_1^i)\right) E(r_j - r_0)}{Eu_i''(W_1^i)} = Cov(W_1^i, r_j)$$

Now define the following measure of agent $i$'s absolute risk aversion:

$$\theta_i := \frac{-Eu_i''(W_1^i)}{Eu_i'(W_1^i)}.$$

Then summing across all agents we have that

$$\begin{aligned} E(r_j - r_0) &= \frac{1}{\sum_{i=1}^K \frac{1}{\theta_i}} Cov(W_1, r_j) \\ &= \frac{1}{\sum_{i=1}^K \frac{1}{\theta_i}} W_0 Cov(r_m, r_j) \end{aligned}$$

where the total wealth at time 1 held in risky assets is $W_1 = \sum_{i=1}^K W_1^i$, $W_0$ is the total wealth in risky assets at time 0, and

$$r_m = \frac{W_1}{W_0} - 1$$

therefore is the return on the market portfolio. Note that this alternative representation tells us more about the risk premium as a function of the aggregate risk aversion across agents in the economy. By linearity we also get that

$$Er_m - r_0 = W_0^M Var(r_m) \frac{1}{\sum_{i=1}^K \frac{1}{\theta_i}},$$

which gives a statement as to the actual magnitude expected excess return on the market portfolio. A high $\theta_i$ corresponds to a high risk aversion and this contributes to making the risk premium larger, as expected. Note that if one agent is very close to being risk neutral then the risk premium (holding that person's initial wealth constant) becomes close to zero. Can you explain why that makes sense?

The derivation of the CAPM when using returns is not completely clear in the sense that finding an equilibrium return does not separate out what is found exogenously and what is found endogenously. One should think of the equilibrium argument as determining the initial price of assets given assumptions on the distribution of the price of the assets at the end of the period. A sketch of how the equilibrium argument would run is as follows:

1. Given the expected value and the covariance of end of period asset prices for all assets

2. Given a utility function for each investor which depends only on mean and variance of end-of-period wealth. Assume that utility decreases as a function of variance and increases as a function of mean. Assume also sufficient differentiability

3. Let investor $i$ have an initial fraction of the total endowment of risky asset $j$.

4. Assume that there is riskless lending and borrowing at a fixed rate $r$. Hence the interest rate is exogenous.

5. Given initial prices of all assets, agent $i$ chooses portfolio weights on risky assets to maximize end of period utility. The difference in price between the initial endowment of risky assets and the chosen portfolio of risky assets is borrowed n/placed in the money market at the riskless rate. (In equilibrium where all assets are being held this implies zero net lending/borrowing.)

6. Compute the solution as a function of the initial prices.

7. Find a set of initial prices such that markets clear, i.e such that the sum of the agents positions in the risky assets sum up to the initial endowment of assets.

8. The prices will reflect characteristics of the agents' utility functions, just as we saw above.

9. Now it is possible to derive the CAPM relation by computing expected returns etc. using the endogenously determined initial prices. This is a purely mathematical exercise translating the formula for prices into formulas involving returns.

Hence CAPM is to be thought of as an equilibrium argument explaining asset prices.

There are of course many unrealistic assumptions underlying the CAPM. The distributional assumptions are clearly problematic. Even if basic securities like stocks were well approximated by normal distributions there is no hope that options would be well approximated due to their truncated payoffs. An answer to this problem is to go to continuous time modelling where 'local normality' holds for very broad classes of distributions but that is outside the scope of this course. Note also that a conclusion of CAPM is that all agents hold the same mixture of risky assets which casual inspection show is not the case. A final problem, originally raised by Roll (1977)[2], concerns the observability of the market portfolio and the logical equivalence between the statement that the market portfolio is efficient and the statement that the CAPM relation holds. To see that observability is a problem think for example of human capital. Economic agents face many decisions over a life time related to human capital - for example whether it is worth taking a loan to complete an education, weighing off leisure against additional work which may increase human capital etc. Many empirical studies use all traded stocks (and perhaps bonds) on an exchange as a proxy for the market portfolio but clearly this is at best an approximation. And what if the test of the CAPM relation is rejected using that portfolio? The relation At the intuitive level, the (9.17) tells us that this is equivalent to the inefficiency of the chosen portfolio. Hence one can always argue that the reason for rejection was not that the model is wrong but that the market portfolio is not chosen correctly (i.e. is not on the portfolio frontier). Therefore, it becomes very hard to truly test the model. While we are not going to elaborate on the enormous literature on testing the CAPM, note also that even at first glance it is not easy to test what is essentially a one period model. To get estimates of the fundamental parameters (variances, covariances, expected returns) one will have to assume that the model repeats itself over time, but when firms change the composition of their balance sheets they also change their betas.

Hence one needs somehow to accommodate betas which change over time and this inevitably requires some statistical compromises.

[2]R. Roll (1977): A critique of the asset pricing theory's test; Part I, Journal of Financial Economics, **4:**pp 129 - 76

# 9.3 Relevant, but not particularly structured, remarks on CAPM

## 9.3.1 Systematic and non-systematic risk

This section follows Huang and Litzenberger's Chapters 3 and 4. We have two versions of the capital asset pricing model. The most "popular" version, where we assumed the existence of a riskless asset whose return is $r_0$, states that the expected return on any asset satisfies

$$Er_i - r_0 = \beta_{i,m}(Er_m - r_0). \tag{9.22}$$

This version we derived in the previous section. The other version is the so-called zero-beta CAPM, which replaces the return on the riskless asset by the expected return on $m's$ zero-covariance portfolio:

$$Er_i - Er_{zm} = \beta_{i,m}(Er_m - Er_{zm}).$$

This version is proved by assuming mean-variance optimizing agents, using that two-fund separation then applies, which means that the market portfolio is on the mean-variance locus (note that we cannot talk about a tangent portfolio in the model with no riskfree asset) and using Proposition 32. Note that both relations state that excess returns (i.e. returns in addition to the riskless returns) are linear functions of $\beta_{im}$.

From now on we will work with the case in which a riskless asset exists, but it is easy to translate to the zero-beta version also. Dropping the expectations (and writing "error terms" instead) we have also seen that if the market portfolio $m$ is efficient, the return on any portfolio (or asset) $q$ satisfies

$$r_q = (1 - \beta_{q,m})r_f + \beta_{q,m}r_m + \epsilon_{q,m}$$

where

$$E\epsilon_{q,m} = E\epsilon_{q,m}r_m = 0.$$

Hence

$$\text{Var}(r_q) = \beta_{q,m}^2 \text{Var}(r_m) + \text{Var}(\epsilon_{q,m}).$$

This decomposes the variance of the return on the portfolio $q$ into its *systematic risk* $\beta_{qm}^2 \text{Var}(r_m)$ and its *non-systematic* or *idiosyncratic* risk $\text{Var}(\epsilon_{q,m})$. The reason behind this terminology is the following: We know that there exists a portfolio which has the same expected return as $q$ but whose variance is $\beta_{qm}^2 \text{Var}(r_m)$ - simply consider the portfolio which invests $1 - \beta_{qm}$ in the riskless asset and $\beta_{q,m}$, in the market portfolio. On the other hand, since this portfolio is efficient, it is clear that we cannot obtain a lower variance if we want an expected return of $Er_q$. Hence this variance is a risk which is correlated with movements in the market portfolio and which is non-diversifiable, i.e. cannot be avoided if we want an expected return of $Er_q$. On the other hand as we have just seen the risk represented by the term $\mathrm{VAR}(\epsilon_{q,m})$ can be avoided simply by choosing a different portfolio which does a better job of diversification without changing expected return.

### 9.3.2 Problems in testing the CAPM

Like any model CAPM builds on simplifying assumptions. The model is popular nonetheless because of its strong conclusions. And it is interesting to try and figure out whether the simplifying assumptions on the behavior of individuals (homogeneous expectations) and on the institutional setup (no taxation, transactions costs) of trading are too unrealistic to give the model empirical relevance. What are some of the obvious problems in testing the model?

First, the model is a one period model. To produce estimates of mean returns and standard deviations, we need to observe years of price data. Can we make sure that the distribution of returns over several years remain the same[3]?

Second (and this a very important problem) what is the 'market portfolio' ? Since investments decisions of firms and individuals in real life are not restricted to stocks and bonds but include such things as real estate, education, insurance, paintings and stamp collections, we should include these assets as well, but prices on these assets are hard to get and some are not traded at all.

A person rejecting the CAPM could always be accused of not having chosen the market portfolio properly. However, note that if 'proper choice' of the market portfolio means choosing an efficient portfolio then this is mathematically equivalent to having the CAPM hold.

This point is the important element in what is sometimes referred to as Roll's critique of the CAPM. When discussing the CAPM it is important to remember which facts are mathematical properties of the portfolio frontier and which are behavioral assumptions. The key behavioral assumption of the CAPM is that the market portfolio is efficient. This assumption gives the CAPM-relation mathematically. Hence it is impossible to separate the claim 'the portfolio $m$ is efficient' from the claim that 'CAPM holds with $m$ acting as market portfolio'.

[3]Mulitiperiod versions exist, but they also face problems with time varying parameters.

### 9.3.3 Testing the efficiency of a given portfolio

Since the question of whether CAPM holds is intimately linked with the question of the efficiency of a certain portfolio it is natural to ask whether it is possible to devise a statistical test of the efficiency of a portfolio with respect to a collection of assets. If we knew expected returns and variances exactly, this would be a purely mathematical exercise. However, in practice parameters need to be estimated and the question then takes a more statistical twist: Given the properties of estimators of means and variances, can we reject at (say) a 5% level that a certain portfolio is efficient? Gibbons, Ross and Shanken (Econometrica 1989, 1121-1152) answer this question - and what follows here is a sketch of their test.

Given a portfolio $m$ and $N$ assets whose excess returns are recorded in $T$ time periods. It is assumed that a sufficiently clear concept of riskless return can be defined so that we can really determine excess returns for each period. NOTE: We will change our notation in this section slightly and assume that $r_p, Er_p$ and $\mu_p$ refer to *excess* returns, mean excess returns and estimated mean excess returns of an asset or portfolio $p$. Hence using this notation the CAPM with a riskless asset will read

$$Er_p = \beta_{p,m} Er_m.$$

We want to test this relation or equivalently whether $m$ is an efficient portfolio in a market consisting of $N$ assets. Consider the following statistical model for the excess returns of the assets given the excess return on the portfolio $m$ :

$$r_{it} = \alpha_i + \gamma_i \, r_{mt} + \epsilon_{it}$$

$$i = 1, \ldots, N \text{ and } t = 1, \ldots, T$$

where $r_{it}$ is the (random) *excess* return[4] of asset $i$ in the $t'$th period, $r_{mt}$ is the observed *excess* return on the portfolio in the $t'$th period, $\alpha_i, \gamma_i$ are constants and the $\epsilon_{it}$'s are normally distributed with $Cov(\epsilon_{it}, \epsilon_{jt}) = \sigma_{ij}$ and $Cov(\epsilon_{it}, \epsilon_{is}) = 0$ for $t \neq s$. Given these data a natural statistical representation of the question of whether the portfolio $m$ is efficient is the hypothesis that $\alpha_1 = \cdots = \alpha_N = 0$. This condition must hold for (9.22) to hold.

To test this is not difficult in principle (but there are some computational tricks involved which we will not discuss here): First compute the MLE's of the parameters. It turns out that in this model this is done merely by computing Ordinary Least Squares estimators for $\alpha, \gamma$ and the covariance

[4] Note this change to excess returns.

matrix for each period $\Sigma$. A so-called Wald test of the hypothesis $\alpha = 0$ can then be performed by considering the test statistic

$$W_0 = \widehat{\alpha} Var(\widehat{\alpha}) \widehat{\alpha}^{-1}$$

which you will learn more about in a course on econometrics. Here we simply note that the test statistic measures a distance of the estimated value of $\alpha$ from the origin. Normally, this type of statistics leads to an asymptotic chi squared test, but in this special model the distribution can be found explicitly and even more interesting from a finance perspective, it is shown in GRS that $W_0$ has the following form

$$W_0 = \frac{(T-N-1)}{N} \frac{\left(\frac{\widehat{\mu}_q^2}{\widehat{\sigma}_q^2} - \frac{\widehat{\mu}_m^2}{\widehat{\sigma}_m^2}\right)}{\left(1 + \frac{\widehat{\mu}_m^2}{\widehat{\sigma}_m^2}\right)}$$

where the symbols require a little explanation: In the minimum variance problem with a riskless asset we found that the excess return of any portfolio satisfies

$$Er_p = \beta_{pm} Er_m.$$

We refer to the quantity

$$\frac{Er_p}{\sigma(r_p)}$$

as the Sharpe ratio for portfolio $p$. The Sharpe ratio in words compares excess return to standard deviation. Note that using the CAPM relation we can write

$$\frac{Er_p}{\sigma(r_p)} = \frac{\sigma(r_m)\rho_{mp}}{\sigma^2(r_m)}(Er_m)$$

where $\rho_{mp}$ is the correlation coefficient between the return of portfolios $p$ and $m$. From this expression we see that the portfolio which maximizes the Sharpe ratio is (proportional) to $m$. Only portfolios with this Sharpe ratio are efficient. Now the test statistic $W_0$ compares two quantities: On one side, the maximal Sharpe ratio that can be obtained when using for parameters in the minimum variance problem the estimated covariance matrix and the estimated mean returns for the economy consisting of the $N$ assets and the portfolio $m$. On the other side, the Sharpe ratio for the particular portfolio $m$ (based on its estimated mean return and standard deviation).

Large values of $W_0$ will reject the hypothesis of efficiency and this corresponds to a case where the portfolio $m$ has a very poor expected return per unit of standard deviation compared to what is obtained by using all assets.