Chapter 3

Risky business: Mean-variance optimal portfolios and the capital asset pricing model

Facts we need

First, the reader should familiarize herself with the Appendix A about probability theory, in particular the concepts of expectation (denoted by E), variance (Var), and covariance (Cov).cSecond, we need some few facts about matrices. (A very useful reference for mathematical results in the large classcimprecisely defined as “well-known” is Berck & Sydsæter (1992), “Economists’ Mathematical Manual”, 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 2×2-matrices.

- Covariance is bilinear. Or more specifically: 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.

	A	B	C	D	E	F	G	H	I	J
1	A cautionary correlation example									
2										
3										
4		Prices			Rates of return			Returns		
5	state/time	Stock 1	Stock 2		Stock 1	Stock 2		Stock 1	Stock 2	
6	0	120	80							
7	1	130	70		0.083333333	-0.125		10	-10	
8	2	140	60		0.076923077	-0.14285714		10	-10	
9	3	150	50		0.071428571	-0.16666667		10	-10	
10										
11		*Correlation*	*-1.000*		*Correlation*	*0.992*		*Correlation*	*#DIV/0!*	
12								*Covariance*	*0*	
13										
14										

Figure 3.1: Uwe Wystup's cautionary correlation example. File: https://tinyurl.com/fe2dc8sn

Example 12. (Inspired by Uwe Wystup.) Ever so often we need to be careful and precise about prices, returns, and rates of return, as well as to whether we mean correlation or covariance. To see why, consider the two-stock example in Figure 12. Someone asks: What is the correlation between the stocks? At that level of specificity the answer can be almost anything. The correlation between prices is -1 (stock 1 goes up-up, stock 2 goes down-down; cell C111), but if we look at rates of return – which is what we use in the mean-variance problem formulation –the correlation is +0.992 (F11), i.e. almost perfectly positive (both rates of return decrease over time because of the numerators; stock 1's become less positive, stock 2's become more negative). To make matters even more confusing: The correlation between returns (price changes) is not well-defined (the standard deviations in the denominator are 0), while the covariance between returns is 0. ■

Basic definitions and 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

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

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 $r = (r_1, \dots, r_n)^\top$ 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 ith 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 coordinates 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

$$\begin{aligned} W_1 &= \sum_{i=1}^{n} a_i S_{i,1} = \sum_{i=1}^{n} a_i (S_{i,1} - S_{i,0}) + \sum_{i=1}^{n} a_i S_{i,0} \\ &= W_0 \left(1 + \sum_{i=1}^{n} \frac{S_{i,0} a_i}{W_0} \frac{S_{i,1} - S_{i,0}}{S_{i,0}} \right) \\ &= W_0 (1 + \mathbf{w}^\top r), \end{aligned} \tag{3.1}$$

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

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

where E denoted expectation (or mean) and

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

In this chapter we will look at how agents should 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 she derives utility 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 conditions this means that the 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 (3.1), 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.

3.1 Mathematics of minimum variance portfolios

3.1.1 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

$$\begin{aligned} E(r) &= \mu, && (3.2) \\ \mathrm{Cov}(r) &= \mathbf{\Sigma}, && (3.3) \end{aligned}$$

where $\mathbf{\Sigma}$ is positive definite (hence invertible) and not all coordinates of μ are equal. As a covariance matrix $\mathbf{\Sigma}$ 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:

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

Analysis of such a problem is called mean/variance analysis, or Markowitz analysis after Harry Markowitz who studied the problem in the 40'ies and 50'ies. (He won the Nobel prize in 1990 together with William Sharpe and Merton Miller both of whom we'll meet later.)
Our assumptions on μ and $\mathbf{\Sigma}$ ensure that a unique finite solution exits for any value of μ_P. 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 portfolio's rate return ($\mathbf{w}^\top \mathbf{\Sigma} \mathbf{w}$; the "1/2" is just there for convenience) is minimized given that we want a specific expected rate of return (μ_P; "P is for portfolio").
To solve the problem we set up the Lagrange-function with multipliers

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

The first-order conditions for optimality are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \mathbf{\Sigma} \mathbf{w} - \lambda_1 \mu - \lambda_2 \mathbf{1} &= 0, && (3.4) \\ \mathbf{w}^\top \mu - \mu_P &= 0, && (3.5) \\ \mathbf{w}^\top \mathbf{1} - 1 &= 0. && (3.6) \end{aligned}$$

Usually we might say "and these are linear equations that can easily be solved", but working on them algebraically leads to a deeper understanding and intuition about the model. Invertibility of $\mathbf{\Sigma}$ gives that we can write (3.4) as (check for yourself)

$$\mathbf{w} = \mathbf{\Sigma}^{-1} [\mu \;\; \mathbf{1}] \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}, \tag{3.7}$$

and (3.5)-(3.6) as

$$[\mu \ \ \mathbf{1}]^\top \mathbf{w} = \begin{bmatrix} \mu_P \\ 1 \end{bmatrix}. \tag{3.8}$$

Multiplying both sides of (3.7) by $[\mu \ \ \mathbf{1}]^\top$ and using (3.8) gives

$$\begin{bmatrix} \mu_P \\ 1 \end{bmatrix} = [\mu \ \mathbf{1}]^\top \mathbf{w} = \underbrace{[\mu \ \mathbf{1}]^\top \mathbf{\Sigma}^{-1} [\mu \ \mathbf{1}]}_{=:\mathbf{A}} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}. \tag{3.9}$$

Using the multiplication rules for partitioned matrices we see that

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

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

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

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

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

because $\mathbf{\Sigma}^{-1}$ is positive definite (because $\mathbf{\Sigma}$ is). In other words, $\mathbf{A}$ is positive definite. Hence we can solve (3.9) for the λ's,

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

and insert this into (3.7) in order to determine the optimal portfolio weights

$$\widehat{\mathbf{w}} = \mathbf{\Sigma}^{-1} [\mu \ \mathbf{1}] \mathbf{A}^{-1} \begin{bmatrix} \mu_P \\ 1 \end{bmatrix}. \tag{3.10}$$

The portfolio $\widehat{\mathbf{w}}$ is called the minimum variance portfolio for a given mean μ_P. (We usually can't be bothered to say the correct full phrase: "minimum variance on rate of return for a given mean rate on return μ_P ".) The minimal portfolio return variance is

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

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

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

we have

$$\widehat{\sigma}_P^2 = \frac{a - 2b\mu_P + c\mu_P^2}{ac - b^2}. \tag{3.11}$$

In (3.11) 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 *minimum variance portfolio frontier* or *locus*.

Note that we have not just solved one "minimize variance" problem, but a whole bunch of them, namely one for each conceivable expected rate of return.

In mean-standard deviation-space the relation is expressed as a hyperbola. Figure 3.2 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 3.2 (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*.

Figure 3.2 also shows the *global minimum variance portfolio*, the portfolio with the smallest possible variance for any given mean return. Its mean, μ_G, is found by minimizing (3.11) with respect to μ_P, and is $\mu_{gmv} = \frac{b}{c}$. By substituting this in the general $\widehat{\sigma}^2$-expression we obtain

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

while the general formula for portfolio weights gives us

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

Example 13. Consider the case with 3 assets (referred to as A, B, and C) and

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

The all-important $\mathbf{A}$-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

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

The locus is illustrated in Figure 3.3 in both in (variance, expected return)-space and (standard deviation, expected return)-space.

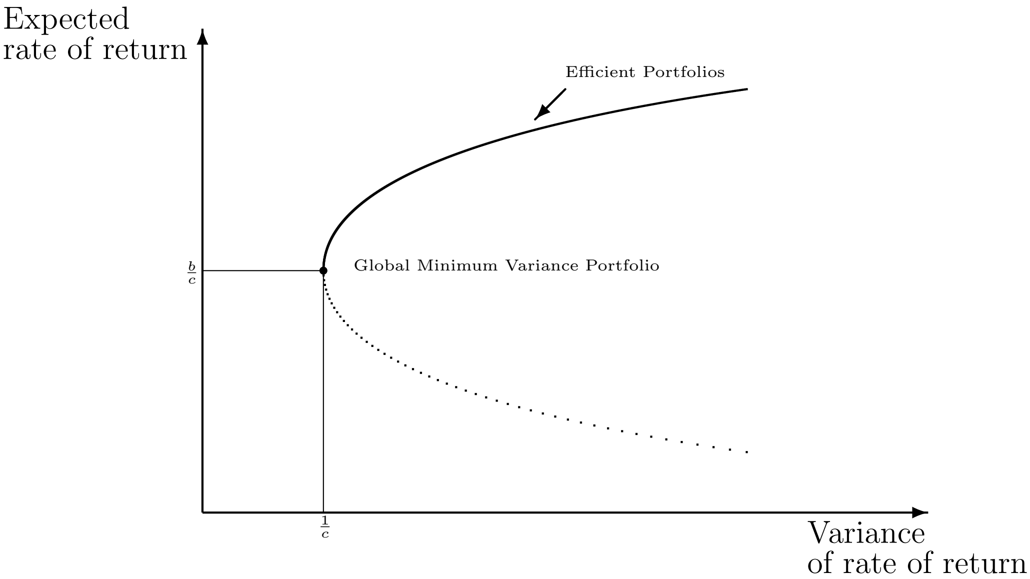

Figure 3.2: The minimum variance portfolio frontier.

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

Proposition 3. *Let* $\mathbf{x}_a$ *and* $\mathbf{x}_b$ *be two minimum variance portfolios with mean returns* μ_a *and* μ_b, $\mu_a \neq \mu_b$. *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 μ_c denote the mean return on a given minimum variance portfolio $\mathbf{x}_c$. Now choose α such that $\mu_c = \alpha\mu_a + (1-\alpha)\mu_b$, that is $\alpha = (\mu_c - \mu_b)/(\mu_a - \mu_b)$ (which is well-defined because $\mu_a \neq \mu_b$). But since $\mathbf{x}_c$ is a minimum variance portfolio we know that (3.10) holds, so

$$\begin{aligned} \mathbf{x}_c &= \mathbf{\Sigma}^{-1}[\mu \ \ \mathbf{1}]\mathbf{A}^{-1} \begin{bmatrix} \mu_c \\ 1 \end{bmatrix} \\ &= \mathbf{\Sigma}^{-1}[\mu \ \ \mathbf{1}]\mathbf{A}^{-1} \begin{bmatrix} \alpha\mu_a + (1-\alpha)\mu_b \\ \alpha + (1-\alpha) \end{bmatrix} \\ &= \alpha\mathbf{x}_a + (1-\alpha)\mathbf{x}_b, \end{aligned}$$

where the third line is obtained because $\mathbf{x}_a$ and $\mathbf{x}_b$ also fulfill (3.10). 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

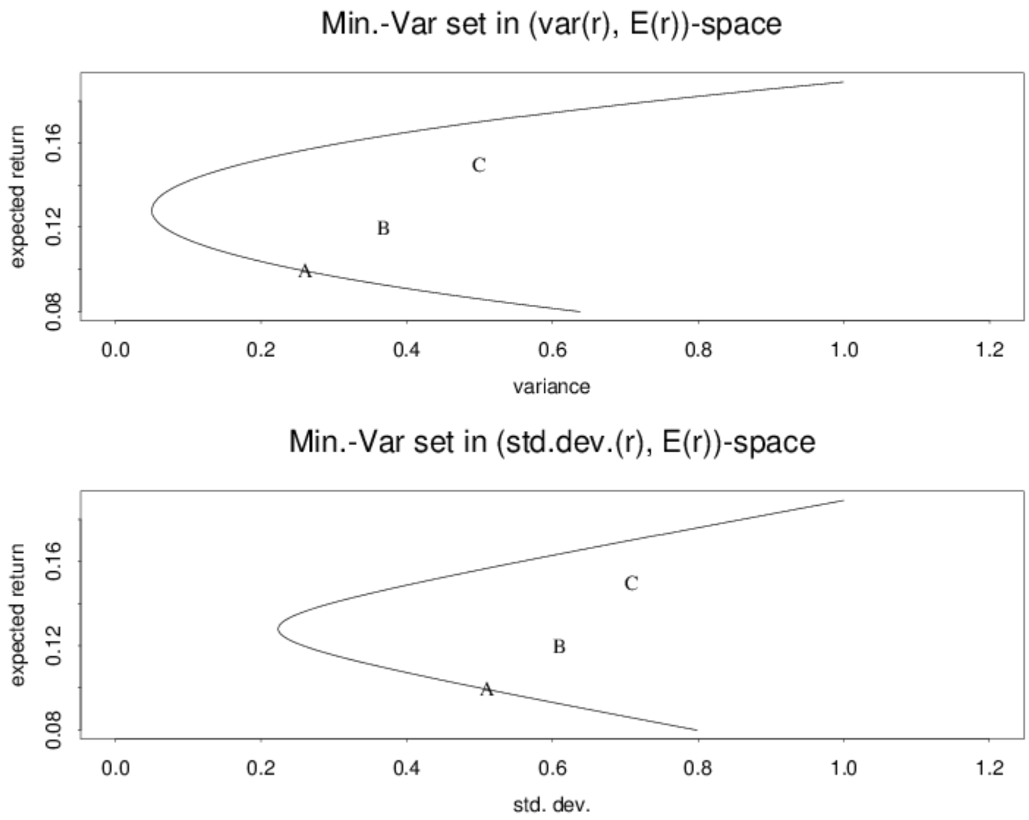

Figure 3.3: The minimum variance frontiers and individual assets

with expected return $\alpha\mu_a + (1-\alpha)\mu_b$. From this, the validity of the third statement is clear. $\square$

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.

$$\mathbf{x}_{zP}^\top \mathbf{\Sigma} \mathbf{x}_P = 0. \tag{3.12}$$

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

Proposition 4. *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 μ_P, its orthogonal one has mean*

$$\mu_{zP} = \frac{a - b\mu_P}{b - c\mu_P}.$$

Proof. First note that μ_{zP} is well-defined for any portfolio except the global minimum variance portfolio. By (3.10) we know how to find the minimum variance portfolios with means μ_P and $\mu_{zP} = (a - b\mu_P)/(b - c\mu_P)$. This leads to

$$\begin{aligned}
\mathbf{x}_{zP}^\top \mathbf{\Sigma} \mathbf{x}_P &= [\mu_{zP}\ \ 1]\mathbf{A}^{-1}[\mu\ \ \mathbf{1}]^\top \mathbf{\Sigma}^{-1}\mathbf{\Sigma}\mathbf{\Sigma}^{-1}[\mu\ \ \mathbf{1}]\mathbf{A}^{-1}[\mu_P\ \ 1]^\top \\
&= [\mu_{zP}\ \ 1]\mathbf{A}^{-1}\underbrace{\left([\mu\ \ \mathbf{1}]^\top \mathbf{\Sigma}^{-1}[\mu\ \ \mathbf{1}]\right)}_{=\mathbf{A}\ \text{ by def.}}\mathbf{A}^{-1}[\mu_P\ \ 1]^\top \\
&= [\mu_{zP}\ \ 1]\mathbf{A}^{-1}\begin{bmatrix}\mu_P \\ 1\end{bmatrix} \qquad (3.13) \\
&= \begin{bmatrix}\frac{a - b\mu_P}{b - c\mu_P} & 1\end{bmatrix}\frac{1}{ac - b^2}\begin{bmatrix} c & -b \\ -b & a\end{bmatrix}\begin{bmatrix}\mu_P \\ 1\end{bmatrix} \\
&= \frac{1}{ac - b^2}\begin{bmatrix}\frac{a - b\mu_P}{b - c\mu_P} & 1\end{bmatrix}\begin{bmatrix} c\mu_P - b \\ a - b\mu_P\end{bmatrix} \\
&= 0,
\end{aligned}$$

which was the desired result. □

Proposition 5. *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 μ_{mv} and variance σ^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

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

*P*roof. Consider first the covariance between return on asset i and $\mathbf{x}_{mv}$. By using (3.10) we get

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

From calculations in the proof of Proposition 4 we know that the covariance between $\mathbf{x}_{mv}$ and $\mathbf{x}_{zvp}$ is given by (3.13). We also know that it is 0. Subtracting this 0 from the above equation gives

$$\begin{aligned}
\mathrm{Cov}(r_i, r_{mv}) &= [\mu_i - \mu_{zmv}\ \ 0]\mathbf{A}^{-1}\begin{bmatrix}\mu_{mv} \\ 1\end{bmatrix} \\
&= (\mu_i - \mu_{zmv})\underbrace{\frac{c\mu_{mv} - b}{ac - b^2}}_{:=\gamma}, \qquad (3.14)
\end{aligned}$$

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 (3.14) we get the desired result for individual assets. But then linearity ensures that it holds for all portfolios. ■

Proposition 5 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.

The converse of Proposition 5 holds in the following sense: If there is a candidate portfolio x_C and a number μ_{zC} such that for any individual asset i we have

$$\mu_i - \mu_{zC} = \beta_{i,C}(\mu_C - \mu_{zC}), \tag{3.15}$$

with $\beta_{i,C} = \mathrm{Cov}(r_i, r_C)/\sigma_C^2$, then x_C is a minimum-variance portfolio. To see why, put $\gamma_i = \sigma_C^2(\mu_i - \mu_{zC})/(\mu_C - \mu_{zC})$ and note that we have $\gamma = \Sigma x_C$, which uniquely determines the candidate portfolio. But by Proposition 5 we know that the minimum variance portfolio with expected rate of return μ_C is (the then) one (and only) portfolio for which (3.15) holds.

3.1.2 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 μ_0. For the risky assets we let ${\mu_i}^e$ denote the *excess* rate of return over the riskfree asset, i.e. the actual rate of return less μ_0. We let μ^e denote the mean excess rate of return, and $\mathbf{\Sigma}$ the covariance matrix (which is of course unaffected). A portfolio is now a $n+1$-dimensional vector whose coordinates 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

$$\mu_P^e = \mathbf{w}^\top \mu^e$$

and the variance is

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

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

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

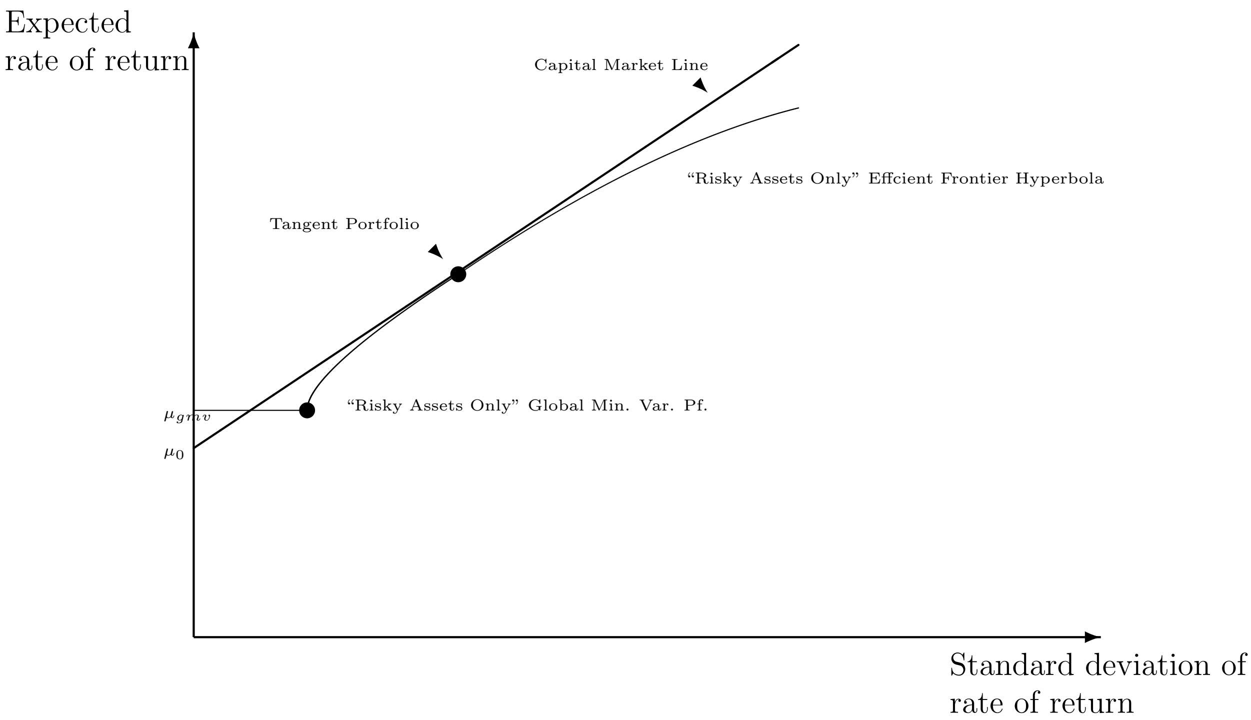

Figure 3.4: The capital market line.

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, and solve the first order conditions. This gives the optimal weights

$$\widehat{\mathbf{w}} = \frac{\mu_P^e}{(\mu^e)^\top \boldsymbol{\Sigma}^{-1} \mu^e} \boldsymbol{\Sigma}^{-1} \mu^e, \tag{3.16}$$

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

$$\widehat{\sigma}_P^2 = \frac{(\mu_P^e)^2}{(\mu^e)^\top \boldsymbol{\Sigma}^{-1} \mu^e}. \tag{3.17}$$

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 $(0, \mu_0)$ with a slope of $\sqrt{(\mu^e)^\top \boldsymbol{\Sigma}^{-1} \mu^e}$. This line is called the capital market line (CML). 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

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

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 3.4, and that explains why we use the term "tangency". When $\mu_{tan}^e > 0$, 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* (after William Sharpe) as excess return relative to standard deviation,

$$\text{Sharpe-ratio}_P = \frac{\mu_P - \mu_0}{\sigma_P}.$$

In the case where $\mu_{tan}^e > 0$, we see from Figure 3.4 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.) The observation that "Higher Sharpe-ratio is better. End of story." makes it a frequently used tool for evaluating and comparing the performance for investment funds. This may sound borderline trivial, but note that if instead we defined Sharpe-ratio with variance in the denominator, then there will be some ineffcient portfolios that have higher Sharpe-ratios that some mean-variance effcient portfolios.
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 5 (and its converse).

Proposition 6. *Let $\mathbf{x}_{mv}$ be a portfolio on the mean-variance frontier with rate of return r_{mv}, expected rate of return μ_{mv} and variance σ_{mv}^2. Let $\mathbf{x}_P$ be an arbitrary portfolio, and use similar notation for rates of return on these portfolios. Then the following holds:*

$$\mu_P - \mu_0 = \beta_{P,mv}(\mu_{mv} - \mu_0),$$

where

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

Conversely, a portfolio for which these equations hold for all individual assets is *on the mean-variance frontier.*

3.1.3 Messing with your head: Effects of parameter uncertainty and the Black-Litterman model

Let $\mathbf{R} = (R_1, R_2, ..., R_N)^T$ be the **excess return** per unit time of N risky assets, i.e. let R_i be defined as the **rate of return** per unit time of the i^{th} asset *minus* the rate of return per unit time of the risk free asset for $i = 1, 2, ..., N$. The mean excess returns per unit time is represented by the vector $\boldsymbol{\mu} = E(\mathbf{R})$, and the covariance matrix per unit time is given by $\boldsymbol{\Sigma} = Var(\mathbf{R}) = E((\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})^T)$. Assume the excess returns to be normally distributed, i.e. $\mathbf{R} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Since the covariance matrix $\boldsymbol{\Sigma}$ by definition is positive definite, it follows that it admits a Cholesky decomposition:

$$\exists \boldsymbol{\sigma} \in \mathbb{R}^{N\times N} \text{ s.t. } \boldsymbol{\Sigma} = \boldsymbol{\sigma}\boldsymbol{\sigma}^T,$$

where $\boldsymbol{\sigma}$ is a lower triangular matrix. In particular, one may readily check that $\mathbf{R}$ thence can be written as

$$\mathbf{R} = \boldsymbol{\mu} + \boldsymbol{\sigma}\mathbf{Z},$$

where $\mathbf{Z}$ is a random vector $\Omega \mapsto \mathbb{R}^n$ with distribution $\sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. More generally, the excess return vector over the time increment Δt is given by

$$\mathbf{R}_{\Delta t} = \boldsymbol{\mu}\Delta t + \boldsymbol{\sigma}\sqrt{\Delta t}\mathbf{Z},$$

in discrete analogy with geometric brownian motion.

Provided that $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known to us, we can compute the mean-variance frontier and the capital market line using standard techniques. Nevertheless, it remains unclear whether these quantities can be reliably estimated, and indeed what bearing a negative answer to this query might have on our capital allocation. To test just how stable the mean-variance frontier is for empirical estimates of the mean and covariance, we design the following experiment:

1. First, to get empirically plausible values for the estimators $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\Sigma}}$ we use daily empirical data for five risky assets based on Kenneth French's "five industry portfolios" and the "Farma-French 3-Factors" available for free at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. Unbiased sample estimates for the components μ_i and Σ_{ij} are given by

$$\bar{\mu}_i := \frac{1}{n\Delta t}\sum_{k=1}^{n}(R_{\Delta t})_{ik}, \tag{3.18}$$

   and

$$\bar{\Sigma}_{ij} := \frac{1}{(n-1)\Delta t}\sum_{k=1}^{n}[(R_{\Delta t})_{ik} - \bar{\mu}_i\Delta t][(R_{\Delta t})_{jk} - \bar{\mu}_j\Delta t], \tag{3.19}$$

   where $\{(R_{\Delta t})_{ij}\}_{j=1}^{n}$ is a time series of n consecutive Δt excess returns for asset i.

2. For our present purposes we shall think of $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\Sigma}}$ as the "true" parameters of the market.

3. Using the equation $\mathbf{R}_{\Delta t} = \bar{\boldsymbol{\mu}}\Delta t + \bar{\boldsymbol{\sigma}}\sqrt{\Delta t}\mathbf{Z}$ where $\boldsymbol{\Sigma} = \bar{\boldsymbol{\sigma}}\bar{\boldsymbol{\sigma}}^T$ we now simulate m future evolutions of the five risky assets over a fixed temporal horizon.

4. For each future evolution in the simulated data, we re-compute sample estimates for the mean and the covariance matrix. Label these by $\hat{\boldsymbol{\mu}}_i$ and $\hat{\boldsymbol{\Sigma}}_i$ where $i = 1, 2, ..., m$. Obviously the *expected values* of the random variables $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ will be the "true" parameters $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\Sigma}}$.

5. Using the mean-variance equation
$$\hat{\sigma}_P^2(\mu_P) = \frac{a - 2b\mu_P + c\mu_P^2}{d}$$
or, equivalently,
$$\mu_p(\hat{\sigma}_P^2) = \frac{b}{c} \pm \sqrt{\frac{1}{c}\left[d\hat{\sigma}_P^2 + \frac{b^2}{c} - a\right]},$$
where $a = \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, $b = \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1} = \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, $c = \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}$, and $d = ac - b^2$ we plot the mean-variance frontier $(\hat{\sigma}_P^2, \mu_P)$ for the scenarios:

   (a) $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ (the simulated sample estimates).
   (b) $(\hat{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}})$ (simulated sample mean, "true" covariance).
   (c) $(\bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ ("true" mean, simulated covariance).

6. We conjecture that the frontiers will be all over the place when we use the simulated sample mean, $\hat{\boldsymbol{\mu}}$. The simple explanation for this is that the real parameter $\bar{\boldsymbol{\mu}}$ is difficult to estimate reliably (as we have seen elsewhere, for log-returns the estimator is a telescoping sum, meaning that only the first and last data point in the stock price process end up determining the expected return). A similar problem does not pertain to the covariance.

The true mean ($\bar{\boldsymbol{\mu}}$) and covariance matrix ($\bar{\boldsymbol{\Sigma}}$) for French's five industry portfolios are exhibited in tables 3.1 and 3.2. The estimators are based on daily data points collected over the 20 year horizon July 1995 to July 2015. The associated mean-variance frontier, market portfolio, and capital market line (CML) are exhibited in the left-hand part of Figure 3.5. Notice that both the frontier and the CML are parabolic functions in (variance, mean)-space - had we plotted the corresponding curves in (standard deviation, mean)-space they would respectively transform to a hyperbola and a straight line. Furthermore, notice the trending inverse relationship between the expected return of the portfolios and their variances (marked by x in the figure): one would probably guess that taking on more risk (volatility) would be compensated by the promises of a higher expected return, but clearly this is not the case in this concrete empirical example!

Out of interest, the righthand side of figure 3.5 also exhibits the mean-variance frontier in the event we enforce the "no short-selling" restriction $\boldsymbol{\pi} \geq 0$.[1] Given the

[1] This problem must be solved numerically; e.g. using MATLAB's quadprog function. For consistency the depicted true mean-variance frontier has also been computed numerically.

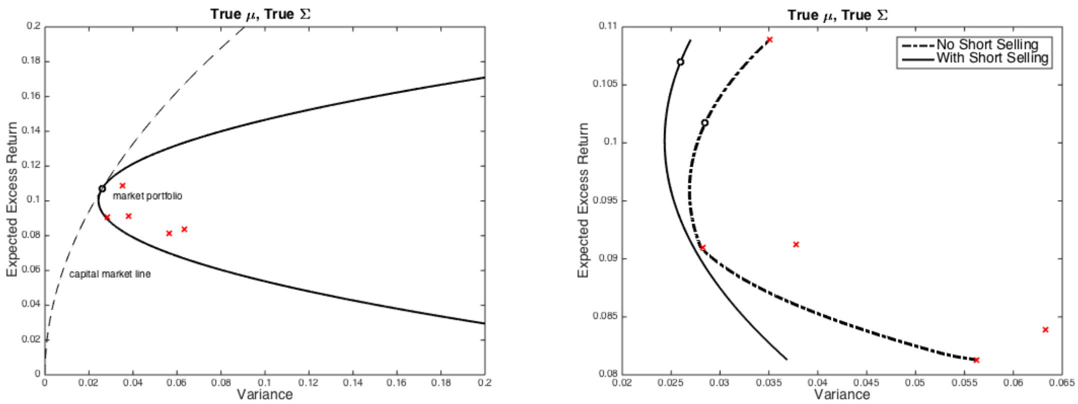

Figure 3.5: **Left:** The mean-variance frontier computed from five industrial indices, marked by x in the diagram. The white dot represents the associated market portfolio, while the dashed line connecting the origin and the market portfolio represents the capital market line. **Right:** The mean-variance frontier with and without short-selling restrictions.

linearity of portfolio returns, such a frontier can only be drawn between the data point with the lowest expected excess return and the data point with the highest expected excess return. Unsurprisingly, a ban on short-selling entails a dampening on the market portfolio from $(\hat{\sigma}_P^2, \mu_P) = (0.0259, 0.1070)$ to $(\hat{\sigma}_P^2, \mu_P) = (0.0285, 0.1017)$, and therefore more conservative Sharpe ratios for rational investors (from 0.6640 to 0.6026).

Asset 1	Asset 2	Asset 3	Asset 4	Asset 5
0.0909	0.0912	0.0839	0.1089	0.0813

Table 3.1: "True" mean excess return for the five industrial indices (French 1995-2015).

Asset 1	Asset 2	Asset 3	Asset 4	Asset 5
0.0282	0.0266	0.0311	0.0236	0.0336
0.0266	0.0378	0.0343	0.0252	0.0373
0.0311	0.0343	0.0633	0.0290	0.0448
0.0236	0.0252	0.0290	0.0351	0.0305
0.0336	0.0373	0.0448	0.0305	0.0563

Table 3.2: "True" covariance matrix for the five industrial indices (French 1995-2015).

Next we simulate 50 evolutions of the five industrial indices five years into the future. The mean-variance frontiers generated by the associated estimators $\hat{\boldsymbol{\mu}}_i$ and $\hat{\boldsymbol{\Sigma}}_i$ for $i = 1, 2, ..., 50$ are exhibited in figure 3.6. Immediately we notice that frontiers which utilise the sample estimator $\hat{\boldsymbol{\mu}}$ are *highly* scattered with respect to the "true" mean-variance frontier, irrespective of whether we use the "true" or the sample covariance. On the other hand, if we use the "true" drift $\bar{\boldsymbol{\mu}}$ the associated frontiers collapse to something resembling the "true" mean-variance frontier. The implication is clear: mis-specifications of the expected mean return of the risky assets will invariably have a devastating impact on the way rational investors think they should invest vis-a-vis how they ought to invest given full information about the governing dynamics. This form of model mis-specification can significantly curb their welfare gains. The problem, of course, is that the "true" drift $\bar{\boldsymbol{\mu}}$ is notoriously unreliable to estimate (recall that for log returns only the first and final data points in the time series go into the estimation). This point is highlighted in the lefthand part of figure 3.7 where we plot the sample parameter $\hat{\boldsymbol{\mu}}$ across the different simulations. Clearly, the simulated drift estimators oscillate wildly around their "true" counterparts. Given the quadratic nature of the covariance estimator, an analogous problem does not prevail here: there is no significant information/welfare loss affecting rational investors in deploying $\hat{\boldsymbol{\Sigma}}$ over $\bar{\boldsymbol{\Sigma}}$.

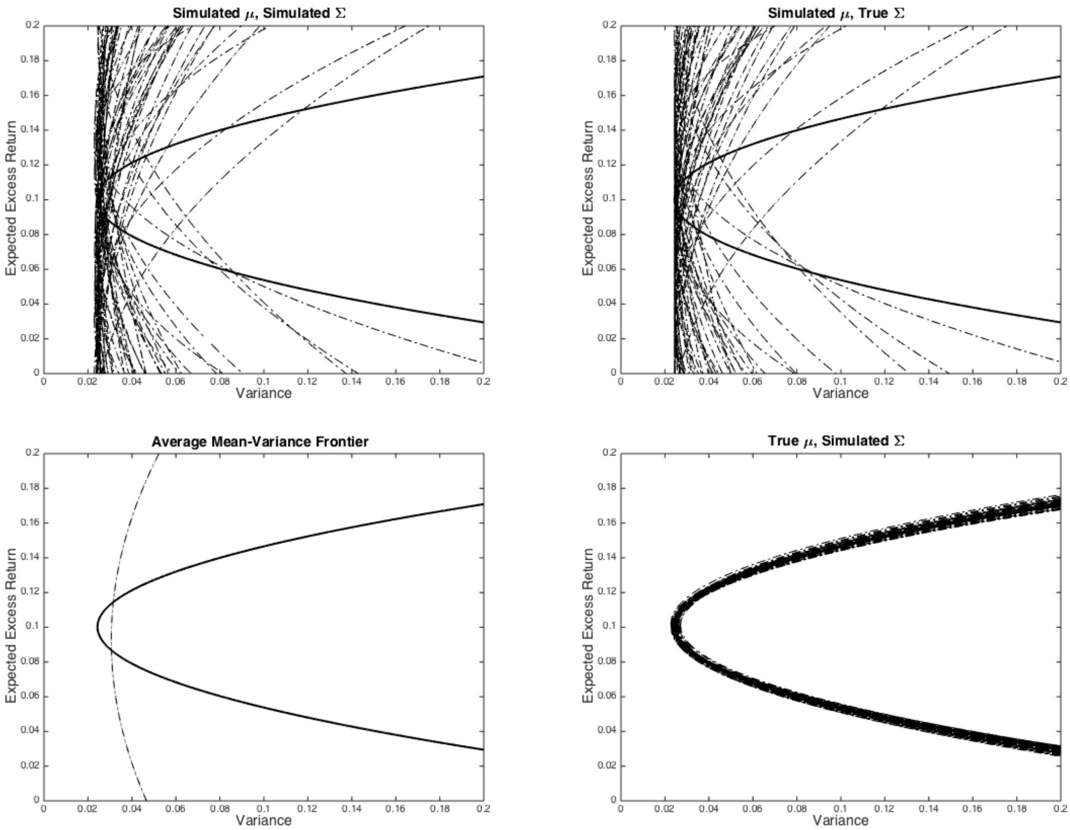

Figure 3.6: **Top left:** The mean-variance frontiers for $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ (the simulated sample estimates). **Top right:** The mean variance-frontiers for $(\hat{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}})$ (simulated sample mean, "true" covariance). **Bottom left:** The average mean-variance frontier for the simulated sample estimates. Specifically, the dashed line represents the horizontal average of the dashed lines in the top left figure. **Bottom right:** The mean variance-frontiers for $(\bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ ("true" mean, simulated covariance).

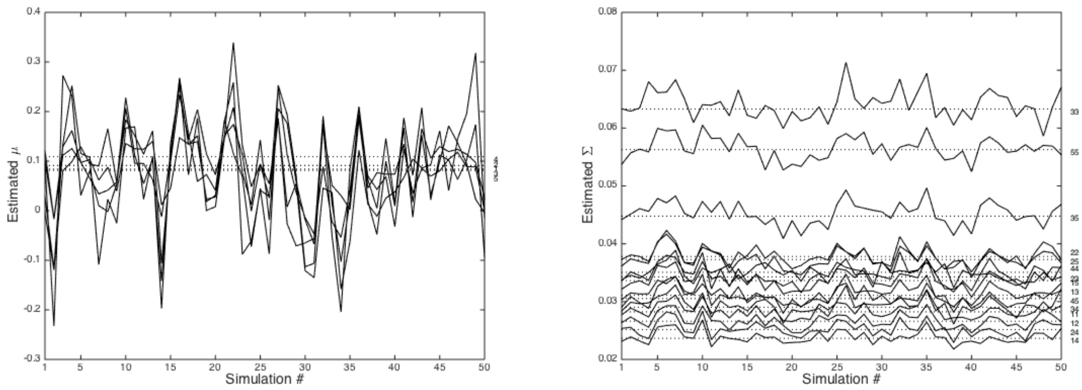

Figure 3.7: **Left:** The stability of the estimator $\hat{\boldsymbol{\mu}}$ across different simulations. The dotted lines represent the "true" means. Clearly, the estimators are highly erratic. **Right:** The stability of the estimator $\hat{\boldsymbol{\Sigma}}$ across different simulations. The dotted lines represent the "true" covariances. Clearly, the estimators are stable.

The Black-Litterman model

Consider an investor solving

$$\max_w w^\top \mu - \frac{\gamma}{2} w^\top \Omega w.$$

Here γ can be interpreted as a (relative) risk-aversion parameter and reasonable values are between 1 and 5. The optimal portfolio weights are

$$\widehat{w} = (\gamma\Omega)^{-1}\mu. \tag{3.20}$$

It is not unreasonable to think of Ω as known. Less so for μ. Imagine now that we reverse the question: Given a postulated risk-aversion and observed portfolio weights, w^{obs}, what should the expected returns be for these weights to be optimal? These we call the equilibrium expected returns, denote by Π, and immediately see that

$$\Pi = \gamma\Omega w^{obs}.$$

A model for noisy equilibrium expected returns is

$$\Pi = I_n\mu + e$$

where $e \sim N(0, \tau\Omega)$ and we call τ the precision.
A model for our views, V, on expected returns as well as our confidence in these, is

$$V = P\mu + v$$

where P is a $k \times n$ (known/specified) matrix and the error $v \sim N(0, \Sigma)$ is assumed independent of e.

As an example, suppose you are very certain in your expectation that stock 1 will outperform stock 3 by 2% on (yearly) average. Then $V_1 = 0.02$, $P_{1,1} = 1$, $P_{1,3} = -1$ and all elements of the first row (and column) of Σ are small.

A model specification combining equilibrium and views is

$$\begin{bmatrix} \Pi \\ V \end{bmatrix} = \begin{bmatrix} I_n \\ P \end{bmatrix} \mu + \epsilon,$$

where $\epsilon \sim N(0, A)$ with

$$A = \begin{bmatrix} \tau\Omega & 0 \\ 0 & \Sigma \end{bmatrix}.$$

After multiplying both sides by $A^{-1/2}$ (i.e. standardizing) we can view this as an ordinary regression model for the unknown expected return μ, which we can then estimate using the usual formula $\hat{\beta} = (X^\top X)^{-1} X^\top y$. Straightforward linear algebra leads to the Black-Litterman formula

$$\widehat{\mu} = [(\tau\Omega)^{-1} + P^\top \Sigma^{-1} P]^{-1} [(\tau\Omega)^{-1}\Pi + P^\top \Sigma^{-1} V]. \tag{3.21}$$

In the original Black-Litterman paper there is more than a touch of mystery about this formula. In practice we would then get portfolio weights consistent with our wievs by plugging $\widehat{\mu}$ from eqn. (3.21) back into (3.20). If we are absolutely certain about our views ($\Sigma = 0$), the formula degenerates to

$$\widehat{\mu} = \Pi + \tau\Omega P^\top (\tau P \Omega P^\top)^{-1} (V - P\Pi). \tag{3.22}$$

3.2 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 μ_i denote the rate of return on the i'th risky asset and we let $\mu_0 = r_0$ denote the riskless rate of return. We assume that μ_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

$$Er_i - \mu_0 = \beta_{i,tan}(E(r_{tan}) - \mu_0) \tag{3.23}$$

where

$$\beta_{i,tan} = \frac{Cov(r_i, r_{tan})}{\sigma^2_{tan}}. \tag{3.24}$$

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

$$w^j = \frac{P_0^j}{\sum_{i=1}^n P_0^i}. \tag{3.25}$$

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 ϕ_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 (3.23). This is the CAPM:

$$E(r_i) - \mu_0 = \beta_{i,m}(E(r_m) - \mu_0), \tag{3.26}$$

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 the rates of return are multivariate normal and that the utility function is twice differentiable and concave:[2]

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

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

$$E\left(u_i'(W_1^i)(r_j - r_0)\right) = 0. \tag{3.27}$$

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

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

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

$$\mathrm{Cov}(g(X), Y) = Eg'(X)\mathrm{Cov}(X,Y)$$

and using this we have the following first order condition:

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

i.e.

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

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

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

[2] This derivation follows Huang and Litzenberger: *Foundations for Financial Economics*. If prices are positive, then returns are bigger than -1, so normality must be an approximation.

Summation over all agents gives us

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

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

$$Er_m - \mu_0 = W_0^M \mathrm{Var}(r_m) \frac{1}{\sum_{i=1}^K \frac{1}{\theta_i}},$$

which gives a statement as to the actual magnitude expected excess return on the market portfolio. A high θ_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. Let the expected value and the covariance of end-of-period asset prices for all assets be given.

2. Suppose further that we are 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 riskfree 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 or placed in the money market at the riskless rate – depending on the sign. (In equilibrium where all assets are being held impliying 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, and thus refrerred to as Roll's cirtique , 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? At the intuitive level, the relation (3.23) 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.

3.3 Remarks on CAPM in no particular order

3.3.1 The single index model

The CAPM says that for any stock or portfolio ($\sim i$) we have

$$\mathrm{E}(r_i) = r_f + \beta \times (\mathrm{E}(r_M) - r_0),$$

where $\beta = \mathrm{cov}(r_i, r_M)/\mathrm{var}(r_M)$, M indicates the market portfolio, and r_f is the risk-free rate. When viewing the expeted rate of return as a function of β, this relation is called the *security market line.*

The sequrity maket line or the CAPM-equation is often expressed in terms of random variables,

$$r_i - r_f = \alpha_i + \beta_i(r_M - r_f) + \epsilon_i, \tag{3.28}$$

where the noise-term ϵ_i is assumed to be independent of the market rate of return r^M. This is called the single-index model and α_i is called Jensen's α. To an empiricist it screams "regression" – in which case people add time-indicies and usually work with excess rates of return, $\widetilde{r}_{i,t} = r_{i,t} - r_{f,t}$.

In the single index model (3.28) we have

$$\mathrm{var}(r_i) = \beta_i^2 \mathrm{var}(r_M) + \mathrm{var}(\epsilon_i).$$

The first term is is called systematic risk, the second term unsystematic or idiosyncratic risk. The reason behind this terminology is the following: We know that there exists a portfolio that has the same expected rate of return as asset i but whose variance is $\beta_i^2 \mathrm{var}\,(r_M)$ – namely the portfolio that has $1 - \beta_i$ in the riskfree asset and β_i, in the market portfolio. On the one hand, because this portfolio is efficient, we cannot obtain a lower variance if we want an expected rate of return of $E(r_i)$. Hence this variance is a risk that is non-diversifiable, i.e. it cannot be avoided if we want an expected rate of return of $E(r_)i$. On the other hand as we have just seen the risk represented by the term $\mathrm{var}(\epsilon_i)$ can be avoided simply by choosing a different portfolio that a better job of diversification without changing expected rate of return.

Note that for the single index model we have

$$\mathrm{cov}(r_i, r_M) = \mathrm{cov}(\beta_i r_M + \epsilon_i, r_M) = \beta_i \mathrm{cov}(r_M, r_M),$$

so the CAPM-form of β_i is still valid, $\beta_i = \text{cov}(r_i, r_M)/\text{var}(r_M)$.

We also see that the single index model has the CAPM as the special case $\alpha_i = 0$ – seemingly a testable restriction. But careful: With muliple assets this must be viewed and tested as a joint hypothesis across assets and Proposition 6 tells us that it is really a restriction or condition on the market (or reference) portfolio M.

Example 14. (An empirical tesing caveat) When testing $\alpha_i = 0$ across assets i, a tempting assumption to make is that all errors terms, ϵ_i's, are uncorrelated. But let's multiply each of the equations (over i) in (3.28) with its corresponding market portfolio weight, say w_i, and sum over the is. This gives

$$\sum_i w_i r_i = \sum_i w_i \alpha_i + r_M \sum_i w_i \beta_i + \sum_i w_i \epsilon_i$$

Now, the left-hand side is just r_M. Let's call the first term on the right hand side $\bar{\alpha}$ and note that the second sum (by definition of β's and the market) is 1. From this we get that

$$\sum_i w_i \epsilon_i = -\bar{\alpha}.$$

So there is a linear combination of the random ϵ_i's that is equal to a non-random constant. Then not only can the ϵ_i's not be uncorrelated, but their covariance matrix must be degenerate. ■

3.3.2 CAPM as a pricing model

Consider a model with n risky assets with expected rate of return vector μ and invertible covariance matrix Σ, and put $\mathbf{1}^\top = (1, \ldots, 1)$. A slight but convenient reparametrization of the search for efficient portfolios is to solve

$$\max_w w^\top \mu - \frac{1}{2}\gamma w^\top \Sigma w \ \text{ s. t. } w^\top \mathbf{1} = 1,$$

for different values of γ, which can be interpreted as a risk-aversion parameter. The optimal portfolios are

$$\widehat{w} = \gamma^{-1}\Sigma^{-1}\left(\mu - \eta(\gamma; \mu, \Sigma)\mathbf{1}\right)$$

where $\eta(\gamma; \mu, \Sigma) = (\mathbf{1}^\top \Sigma^{-1}\mu - \gamma)/\mathbf{1}^\top \Sigma^{-1}\mathbf{1}$ can be interpreted as the expected rate of return on $\widehat{w}$'s zero-beta portfolio.

It seems intuitively reasonable that $\partial \widehat{w}_i / \partial \mu_i > 0$, meaning that if an asset's expected rate of return goes up, then so does its weight in any efficient portfolio. Assets for which this does not hold, we could call financial Giffen goods. We will now show that in the mean/variance optimization setting, there are no financial Giffen goods. To do this we look at the problem with the modified expected return vector $\mu + \alpha e_i$, where $\alpha \in \mathbb{R}$ and e_i is the i'th unit vector. The optimal portfolio in this case we can write as

$$\widehat{w}(\alpha) = \widehat{w} + \alpha h,$$

where $h = \gamma^{-1}(\Sigma^{-1} e_i - \frac{e_i^\top \Sigma^{-1} \mathbf{1}}{\mathbf{1}^\top \Sigma^{-1} \mathbf{1}} \Sigma^{-1} \mathbf{1})$. Showing that $\partial \widehat{w}_i / \partial \mu_i > 0$ amounts to proving positivity of the i'th coordinate of h, which we can write as

$$e_i^\top h = \gamma^{-1} \left(e_i^\top \Sigma^{-1} e_i - \frac{(e_i^\top \Sigma^{-1} \mathbf{1})^2}{\mathbf{1}^\top \Sigma^{-1} \mathbf{1}} \right).$$

Because Σ^{-1} is strictly positive definite and symmetric, $x^\top \Sigma^{-1} y$ defines an inner product, and strict positivity of the term in parenthesis on the right hand side of the equation above follows immediately from the Cauchy-Schwartz inequality.

The inclusion of a risk-free asset is handled in the same way with η replaced by the risk-free rate of return because the risk-free asset is any portfolio's zero-beta portfolio.

With this result a newly introduced $(n + 1)$'st asset (or "project") will be in positive demand (or: "attractive") precisely if there is strict inequality in the CAPM-like expression

$$\mu_{n+1} - r > \frac{\mathrm{cov}(r_{n+1}, r_M)}{\mathrm{var}(r_M)} (\mu_M - r), \tag{3.29}$$

where M denotes the market (or tangent) portfolio, and rs with subscripts are (stochastic) rates of returns. We know from Proposition 6 that w is mean-variance efficient precisely if for any individual asset i we have

$$\mu_i - r = \frac{\mathrm{cov}(r_i, r_w)}{\mathrm{var}(r_w)} (\mu_w - r).$$

For the portfolio $(w_M^\top, 0)^\top$ the n first necessary equations hold because the market portfolio is efficient in the old economy, and we see that the new asset is in 0-demand if equality holds in (3.29). Now the absence of Giffen tells us that if there is strict inequality as stated, the $(n + 1)$'st asset has strictly positive weight in the new market portfolio.
This gives us a theoretically well-founded way to evaluate projects in the way illustrated by the following example.

Example 15. Consider a setting where

$$\mathrm{E}(r_M) = 0.07, \quad \sigma_M := \sqrt{\mathrm{var}(r_M)} = 0.15 \text{ and } r_0 = 0.04.$$

Suppose we are given the opportunity (at time 0) to invest in a project that pays at time 1 the stochastic amount X_1 about which we know that

$$\mathrm{E}(X_1) = 1000, \quad \sigma_X := \sqrt{\mathrm{var}(X_1)} = 400, \text{ and } \rho_{M,X} := \mathrm{corr}(r_M, X_1) = 0.5.$$

An interpretation could be: The project is an oil field and our future income depends partly on how much oil is there, partly on the price at which we can sell it in the market. Only the latter is correlated with the (general) financial market, the oil is either there or it isn't. So what is the CAPM-critical price at time 0, say X_0? Written out in detail, the CAPM says

$$\frac{\mathrm{E}(X_1) - X_0}{X_0} = r_0 + \frac{\operatorname{cov}(\frac{X_1 - X_0}{X_0}, r_M)}{\sigma_M^2}(\mathrm{E}(r_M) - r_0), \tag{3.30}$$

which we have to solve for the critical time 0-price X_0^*. To do this let us look at

$$\begin{aligned}\operatorname{cov}\left(\frac{X_1 - X_0}{X_0}, r_M\right) &= \frac{1}{X_0}\operatorname{cov}(X_1, r_M) \\ &= \frac{1}{X_0}\sigma_X \sigma_M \rho_{M,X} \\ &= \frac{400 \cdot 0.15 \cdot 0.5}{X_0} = \frac{30}{X_0}.\end{aligned}$$

We now rearrange equation (3.30) to get

$$\underbrace{\mathrm{E}(X_1)}_{1000} = X_0 \underbrace{(1 + r_0)}_{1.04} + \underbrace{\frac{30}{\sigma_M^2}(\mathrm{E}(r_M) - r_0)}_{40}$$

which leads to

$$X_0^* = \frac{960}{1.04} = 923.1.$$

So if the project cost less that 923.1, it is attractive, otherwise not. The equilibrium expected rate of return on the project, the left hand side of (3.30) evaluated at X_0^* is 0.0832, and the equilibrium beta of the project is $\beta_X = 30/(X_0^* \sigma_M^2) = 1.44$. Note the quantitative caveat: beta (1.44 here) is not correlation (0.5 here). ■

3.3.3 Messing with your head: How security market lines actually look

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?

Second (and this a very important problem), what is the market portfolio? Since investment 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'.

According to the capital asset pricing model, CAPM, the expected excess return of a risky asset is proportional to the expected excess return of the market portfolio. The constant of proportionality (the so-called beta) is given by the covariance between the (rate of) return of that risky asset and the (rate of) return on the market portfolio *scaled* such that beta of the market portfolio itself is unity:

$$\beta_i = \frac{Cov(r_i, r_m)}{Var(r_m)}.$$

Thus, we may construe the β of an asset as encoding its susceptibility to systemic risk: the higher the beta the higher the co-movement with the market portfolio (and accordingly also the more prone the asset will be to plummet insofar as the market crashes).

In a perfect market, if we were to plot the excess return of all risky assets against their beta, they would form a straight line in the $(\beta, E[r_i] - r_f)$ diagram, with a slope of $E[r_m] - r_f$ and an ordinate intercept of zero. This is the so-called security market line (SML). Any asset falling above it is said to have positive (Jensen) alpha and is under-valued with respect to its level of uncertainty; conversely, any asset falling below the SML is said to have negative alpha and is over-valued.

To test the empirical verisimilitude of this model we have in Figure 3.8 plotted the excess return of 49 industrial portfolios against their beta. The data for the associated linear regression model

$$\text{excess return}_i = \text{slope} \cdot \text{beta}_i + \text{intercept} + \text{error term}_i,$$

is exhibited in table 3.3. Immediately we notice that the slope of the empirical sequrity market line is much flatter than the one predicted by the theoretical CAPM (gradient 0.0128 versus 0.0808). This has the somewhat surprising implication that taking on more systemic risk (higher β) is not "sufficiently compensated" in terms of added expected excess return. Low β stocks therefore seem more appetising.

	Estimate	SE	t-Stat	p-value
Intercept:	0.080798	6.3174e-10	1.279e+08	3.3455e-114
Slope:	0.012853	6.7344e-10	1.9086e+07	8.2528e-102

Table 3.3: Linear regression data of 49 industrial indices (French 1995-2015).

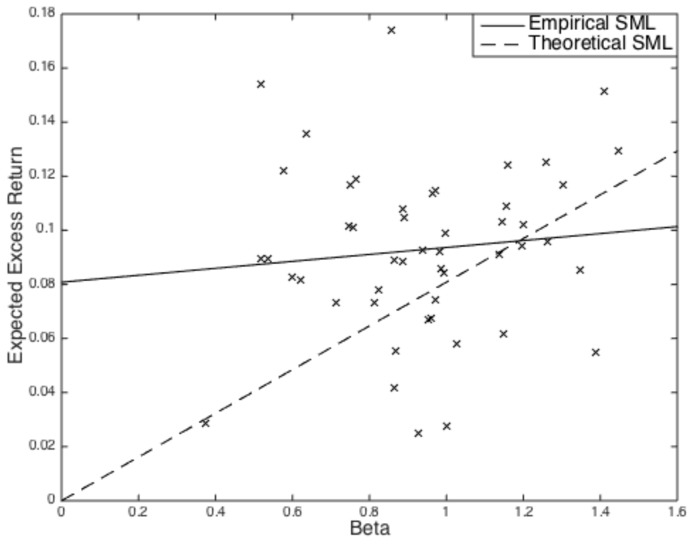

Figure 3.8: The empirical security market line computed on behalf of 49 industrial portfolios (French 1995-2015). On the same graph the theoretical "CAPM" counterpart has been plotted.