Lecture Notes in Finance 1 (MiQE/F, MSc course at UNISG)

Paul Söderlind[1]

11 December 2013

[1]University of St. Gallen. *Address:* s/bf-HSG, Rosenbergstrasse 52, CH-9000 St. Gallen, Switzerland. *E-mail:* Paul.Soderlind@unisg.ch. Document name: Fin1MiQEFAll.TeX

Contents

1 The Basics of Portfolio Choice

Reference: Elton, Gruber, Brown, and Goetzmann (2010) 4–6; Fabozzi, Focardi, and Kolm (2006) 4

1.1 Portfolio Return: Definition, Mean and Variance

Many portfolio choice models center around two moments of the chosen portfolio: the expected return and the variance. This section is therefore devoted to discussing how these moments of the portfolio return are related to the corresponding moments of the underlying assets.

1.1.1 Portfolio Return: Definition

The *net return* on asset i in period t is

$$R_{i,t} = \frac{\text{Value}_{i,t} - \text{Value}_{i,t-1}}{\text{Value}_{i,t-1}} = \frac{\text{Value}_{i.t}}{\text{Value}_{i,t-1}} - 1. \tag{1.1}$$

The gross return is

$$1 + R_{i,t} = \frac{\text{Value}_{i,t}}{\text{Value}_{i,t-1}}. \tag{1.2}$$

Example 1.1 *(Returns)*

$$\begin{aligned} R &= \frac{110 - 100}{100} = 0.1 \textit{ (or } 10\%\textit{)} \\ 1 + R &= \frac{110}{100} = 1.1 \end{aligned}$$

In many cases, the values are

$$\begin{aligned} \text{Value}_{i,t-1} &= P_{i,t-1} \text{ (price yesterday)} \\ \text{Value}_{i,t} &= D_{i,t} + P_{i,t} \text{ (dividend + price today)}, \end{aligned} \tag{1.3}$$

so the return can be written

$$
\begin{aligned}
R_{i,t} &= \frac{D_{i,t} + P_{i,t} - P_{i,t-1}}{P_{i,t-1}} \\
&= \underbrace{\frac{D_{i,t}}{P_{i,t-1}}}_{\text{dividend yield}} + \underbrace{\frac{P_{i,t} - P_{i,t-1}}{P_{i,t-1}}}_{\text{capital gain yield}}
\end{aligned} \tag{1.4}
$$

Example 1.2 *(Dividend yield ad capital gain yield)*

$$
R = \frac{2}{100} + \frac{108 - 100}{100} = 0.1
$$

Let $R_{i,t}$ denote the return on asset i over a given time period. The return on a portfolio ($R_{p,t}$) with the portfolio weights $w_1, w_2, ..., w_n$ ($\Sigma_{i=1}^{n} w_i = 1$) is

$$
\begin{aligned}
R_{p,t} &= w_1 R_{1,t} + w_2 R_{2,t} \text{ (with } n = 2) & (1.5) \\
&= \sum_{i=1}^{n} w_i R_{i,t} \text{ (more generally).} & (1.6)
\end{aligned}
$$

Proof. (of (1.6)) Suppose we bought the number θ_i of asset i in period $t-1$. The total cost of the portfolio was therefore $W_{t-1} = \sum_{i=1}^{n} \theta_i P_{i,t-1}$, where $P_{i,t-1}$ denotes the price of asset i in period $t-1$. Define the portfolio weights as

$$
w_i = \frac{\theta_i P_{i,t-1}}{W_{t-1}}.
$$

The value in period t is $W_t = \sum_{i=1}^{n} \theta_i (D_{i,t} + P_{i,t})$, which we can rewrite (using $\theta_i = w_i W_{t-1}/P_{i,t-1}$) as

$$
W_t = \textstyle\sum_{i=1}^{n} \underbrace{\frac{W_{t-1} w_i}{P_{i,t-1}}}_{\theta_i} (D_{i,t} + P_{i,t}) = W_{t-1} \textstyle\sum_{i=1}^{n} w_i \underbrace{\frac{D_{i,t} + P_{i,t}}{P_{i,t-1}}}_{1+R_{i,t}}.
$$

Divide by W_{t-1} to get the gross The portfolio return

$$
\frac{W_t}{W_{t-1}} = \textstyle\sum_{i=1}^{n} w_i (1 + R_{i,t}) = 1 + \sum_{i=1}^{n} w_i R_{i,t},
$$

where the last equality follows from $\sum_{i=1}^{n} w_i = 1$. Subtract 1 from both sides to get the net portfolio return (1.6). ∎

Example 1.3 *(Number of assets and portfolio returns) For asset* 1 *we have* $P_{1,t-1} = 10$, $P_{1,t} = 11$ *and for asset* 2 $P_{2,t-1} = 8$, $P_{2,t} = 8.4$. *There are no dividends. Yesterday you bought* 16 *of asset* 1 *and* 5 *of asset* 2: $16 \times 10 + 5 \times 8 = 200$. *Today your portfolio is worth* $16 \times 11 + 5 \times 8.4 = 218$, *so* $R_p = \frac{218-200}{200} = 9\%$. *Compare that to (1.6) which would give*

$$R_p = 0.8 \times 10\% + 0.2 \times 5\% = 9\%,$$

since the two returns are 10% *(*$11/10 - 1$*) and* 5% *(*$8.4/8 - 1$*) respectively, and the portfolio weights are* 0.8 *(*$16 \times 10/200$*) and* 0.2 *(*$5 \times 8/200$*) respectively.*

1.1.2 Portfolio Return: Expected Value and Variance

Remark 1.4 *(Expected value and variance of a linear combination) Recall that*

$$\begin{aligned}
\mathrm{E}(aR_1 + bR_2) &= a \,\mathrm{E}\, R_1 + b \,\mathrm{E}\, R_2, \text{ and} \\
\mathrm{Var}(aR_1 + bR_2) &= a^2\sigma_{11} + b^2\sigma_{22} + 2ab\sigma_{12},
\end{aligned}$$

where $\sigma_{ij} = \mathrm{Cov}(R_i, R_j)$, *and* $\sigma_{ii} = \mathrm{Cov}(R_i, R_i) = \mathrm{Var}(R_i)$.

Remark 1.5 *(On the notation in these lecture notes*) Mean returns are denoted* $\mathrm{E}\, R_i$ *or* μ_i. *An expression like* $\mathrm{E}\, R_i^2$ *means the expected value of* R_i^2 *similar to* $\mathrm{E}(R_i^2)$ *and* $\mathrm{E}\, xy$ *is the expectation of the product* xy. *Variances are denoted* σ_i^2 *and sometimes* $\mathrm{Var}(R_i)$ *and the standard deviations* σ_i *or* $\mathrm{Std}(R_i)$. *Covariances are denoted* σ_{ij} *or sometimes* $\mathrm{Cov}(R_i, R_i)$. *Clearly, the covariance* σ_{ii} *must be the same as the variance* σ_i^2.

The expected return on the portfolio is (time subscripts are suppressed to save ink)

$$\begin{aligned}
\mathrm{E}\, R_p &= w_1 \,\mathrm{E}\, R_1 + w_2 \,\mathrm{E}\, R_2 \text{ (with } n = 2) && (1.7) \\
&= \sum_{i=1}^{n} w_i \,\mathrm{E}\, R_i \text{ (more generally).} && (1.8)
\end{aligned}$$

Let $\sigma_{ij} = \mathrm{Cov}(R_i, R_j)$, and $\sigma_{ii} = \mathrm{Cov}(R_i, R_i) = \mathrm{Var}(R_i)$. The variance of a portfolio return is then

$$\begin{aligned}
\mathrm{Var}(R_p) &= w_1^2\sigma_{11} + w_2^2\sigma_{22} + 2w_1w_2\sigma_{12} \text{ (with } n = 2) && (1.9) \\
&= \sum_{i=1}^{n} w_i^2\sigma_{ii} + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} w_i w_j \sigma_{ij} \text{ (more generally).} && (1.10)
\end{aligned}$$

In matrix form we have

$$\mathrm{E}\,R_p = w'\,\mathrm{E}\,R \text{ and} \tag{1.11}$$

$$\mathrm{Var}(R_p) = w'\Sigma w. \tag{1.12}$$

Remark 1.6 *(Details on the matrix form) With two assets, we have the following:*

$$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \mathrm{E}\,R = \begin{bmatrix} \mathrm{E}\,R_1 \\ \mathrm{E}\,R_2 \end{bmatrix}, \textit{ and } \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}.$$

$$\begin{aligned} \mathrm{E}\,R_p &= w'\,\mathrm{E}\,R \\ &= \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} \mathrm{E}\,R_1 \\ \mathrm{E}\,R_2 \end{bmatrix} \\ &= w_1\,\mathrm{E}\,R_1 + w_2\,\mathrm{E}\,R_2. \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(R_p) &= w'\Sigma w \\ &= \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\ &= \begin{bmatrix} w_1\sigma_{11} + w_2\sigma_{12} & w_1\sigma_{12} + w_2\sigma_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\ &= w_1^2\sigma_{11} + w_2 w_1 \sigma_{12} + w_1 w_2 \sigma_{12} + w_2^2 \sigma_{22}. \end{aligned}$$

1.2 The Effect of Diversification

Consider an equally weighted (EW) portfolio of two risky assets. Use $w_1 = w_2 = 1/2$ in (1.9) and assume (for illustrative purposes) that both assets have the same variance (σ^2) and a correlation of ρ. We then get (since $\sigma_{12} = \rho\sqrt{\sigma_{11}\sigma_{22}}$)

$$\mathrm{Var}(R_p) = \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 + \frac{2}{4}\rho\sigma^2 = \frac{1}{2}\sigma^2(1+\rho). \tag{1.13}$$

If the assets are uncorrelated ($\rho = 0$), then this portfolio variance is half the asset variance—which demonstrates the importance of diversification. This effect is event stronger when the correlation becomes negative: with $\rho = -1$ the portfolio variance

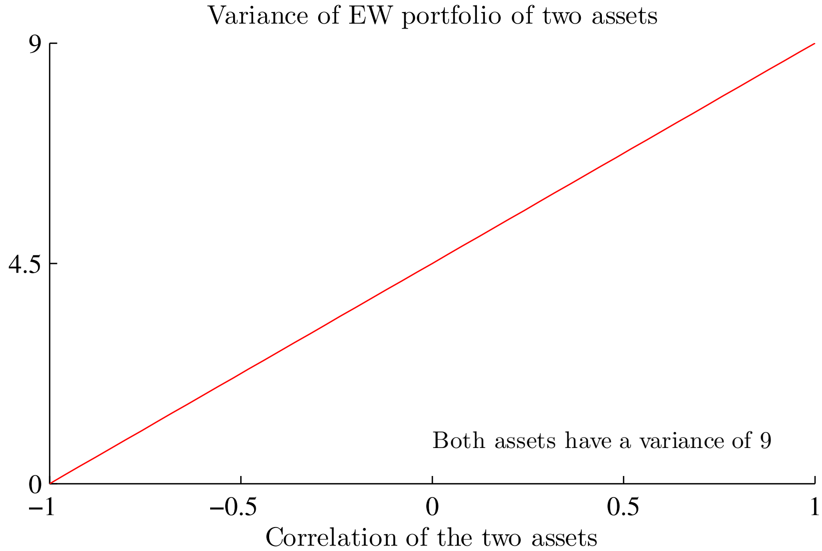


Figure 1.1: Effect of correlation on the diversification benefits

is actually zero (hedging). In contrast, with a high correlation, the benefit from diversification is much smaller (and zero when the correlation is perfect, $\rho = 1$). See Figure 1.1 for an illustration.

In order to see the importance of mixing many assets in the portfolio, start by assuming that the returns are uncorrelated ($\sigma_{ij} = 0$ if $i \neq j$). This is clearly not realistic, but provides a good starting point for illustrating the effect of diversification. We will consider equally weighted portfolios of n assets ($w_i = 1/n$). There are other portfolios with lower variance (and the same expected return), but it provides a simple analytical case.

The variance of an equally weighted ($w_i = 1/n$) portfolio is (when all covariances are zero)

$$\mathrm{Var}(R_p) = \sum_{i=1}^{n} \overbrace{\frac{1}{n^2}}^{w_i^2} \sigma_{ii} = \frac{1}{n} \sum_{i=1}^{n} \frac{\sigma_{ii}}{n} \tag{1.14}$$

$$= \frac{1}{n} \overline{\sigma}_{ii}, \text{ (if } \sigma_{ij} = 0). \tag{1.15}$$

In this expression, $\overline{\sigma}_{ii}$ is the average variance of an individual return. This number could be treated as a constant (that is, not depend on n) if we form portfolios by randomly picking assets. In any case, (1.15) shows that the portfolio variance goes to zero as the number of assets (included in the portfolio) goes to infinity. Also a portfolio with a large but finite number of assets will typically have a low variance (unless we have systematically picked the very most volatile assets).

Second, we now allow for correlations of the returns. The variance of the equally weighted portfolio is then

$$\operatorname{Var}(R_p) = \frac{1}{n}\left(\overline{\sigma}_{ii} - \overline{\sigma}_{ij}\right) + \overline{\sigma}_{ij}, \tag{1.16}$$

where $\overline{\sigma}_{ij}$ is the average covariance of two returns (which, again, can be treated as a constant if we pick assets randomly). Realistically, $\overline{\sigma}_{ij}$ is positive. When the portfolio includes many assets, then the average covariance dominates. In the limit (as n goes to infinity), only this non-diversifiable risk matters.

See Figure 1.2 for an example.

Proof. (of (1.16)) The portfolio variance is

$$\begin{aligned}
\operatorname{Var}(R_p) &= \sum_{i=1}^{n} \frac{1}{n^2}\sigma_{ii} + \sum_{i=1}^{n} \sum_{j=1,,j\neq i}^{n} \frac{1}{n^2}\sigma_{ij} \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{\sigma_{ii}}{n} + \frac{n-1}{n}\sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} \frac{\sigma_{ij}}{n\,(n-1)} \\
&= \frac{1}{n}\overline{\sigma}_{ii} + \frac{n-1}{n}\overline{\sigma}_{ij},
\end{aligned}$$

which can be rearranged as (1.16). ■

A (NoDur)
B (Durbl)
C (Manuf)
D (Enrgy)
E (HiTec)
F (Telcm)
G (Shops)
H (Hlth)
I (Utils)
J (Other)

Table 1.1: Industries

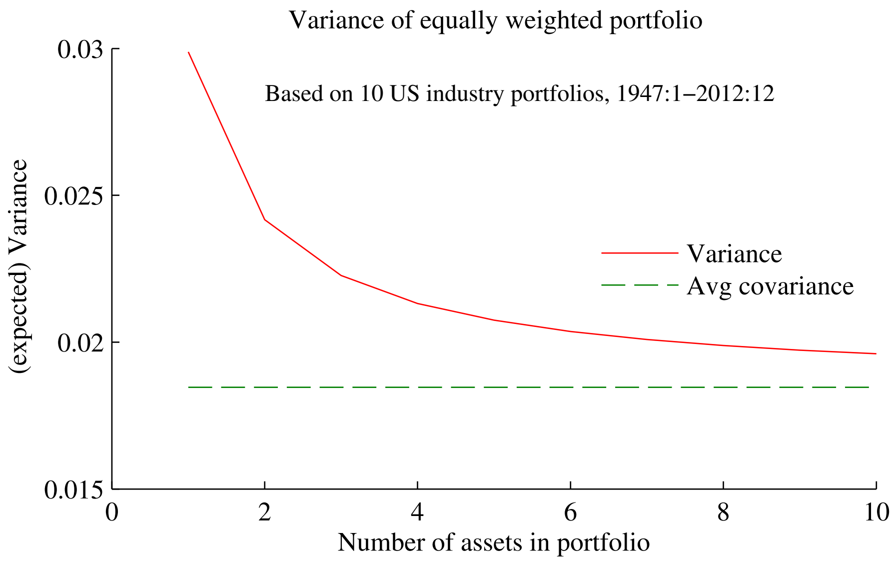


Figure 1.2: Effect of diversification

Remark 1.7 *(On negative covariances in (1.16)*) Formally, it can be shown that $\overline{\sigma}_{ij}$ must be non-negative as $n \to \infty$. It is simply not possible to construct a very large number of random variables (asset returns or whatever other random variable) that are, on average, negatively correlated with each other. In (1.16) this manifests itself in that $\overline{\sigma}_{ij} < 0$ would give a negative portfolio variance as n increases.*

1.2.1 Some Practical Remarks: Annualizing Means and Variances

Remark 1.8 *(Annualizing the MV figures*) Suppose we have weekly net returns $R_t = P_t/P_{t-1} - 1$. The standard way of annualizing the mean and the standard deviation is to first estimate means and the covariance matrix on weekly returns, do all the MV calculations, and then (when showing the results) multiply the mean weekly return by 52 and the standard deviation of the weekly return by $\sqrt{52}$. To see why, notice that an annual return would be*

$$\begin{aligned}
P_t/P_{t-52} - 1 &= (P_t/P_{t-1})(P_{t-1}/P_{t-2})\dots(P_{t-51}/P_{t-52}) - 1 \\
&= (R_t + 1)(R_{t-1} + 1)\dots(R_{t-51} + 1) - 1 \\
&\approx R_t + R_{t-1} + \dots + R_{t-51}.
\end{aligned}$$

To a first approximation, the mean annual return would therefore be

$$\mathrm{E}(R_t + R_{t-1} + \ldots + R_{t-51}) = 52\,\mathrm{E}\,R_t,$$

and if returns are iid (in particular, same variance and uncorrelated across time)

$$\begin{aligned}\mathrm{Var}(R_t + R_{t-1} + \ldots + R_{t-51}) &= 52\,\mathrm{Var}(R_t) \Rightarrow \\ \mathrm{Std}(R_t + R_{t-1} + \ldots + R_{t-51}) &= \sqrt{52}\,\mathrm{Std}(R_t).\end{aligned}$$

1.3 Portfolio Choice: A Risky Asset and a Riskfree Asset

How much to put into the risky asset is a matter of *leverage*.

We typically define the leverage ratio as the investment (into risky assets) divided by how much capital we own

$$\text{Leverage ratio } (v) = \frac{\text{investment into risky assets}}{\text{own capital}}, \tag{1.17}$$

which here equals v. To see the effect on the mean and the volatility of the leverage notice that

$$\begin{aligned} R_p &= vR_1 + (1-v)R_f, \text{ so} \\ \mathrm{E}\,R_p &= v\,\mathrm{E}\,R_1 + (1-v)R_f \text{ and} \qquad (1.18)\\ \mathrm{Std}(R_p) &= |v|\,\mathrm{Std}(R_1). \qquad (1.19)\end{aligned}$$

Both the mean and the standard deviation are scaled by the leverage ratio. Figure 1.3 illustrates the effect on the portfolio return distribution.

As long as the leverage ration is positive ($v > 0$), we can combine these equations to get

$$\mathrm{E}\,R_p = R_f + \mathrm{Std}(R_p) \times SR_1, \tag{1.20}$$

where $SR_1 = \left(\mathrm{E}\,R_1 - R_f\right) / \,\mathrm{Std}(R_1)$ is the *Sharpe ratio* of the risky (first) asset. This shows that the average portfolio return is linearly related to its standard deviation. See Figure 1.3.

Suppose now that the investor seeks to trade off expected return and the variance of the portfolio return. In the simplest case of one risky asset (stock market index, say) and

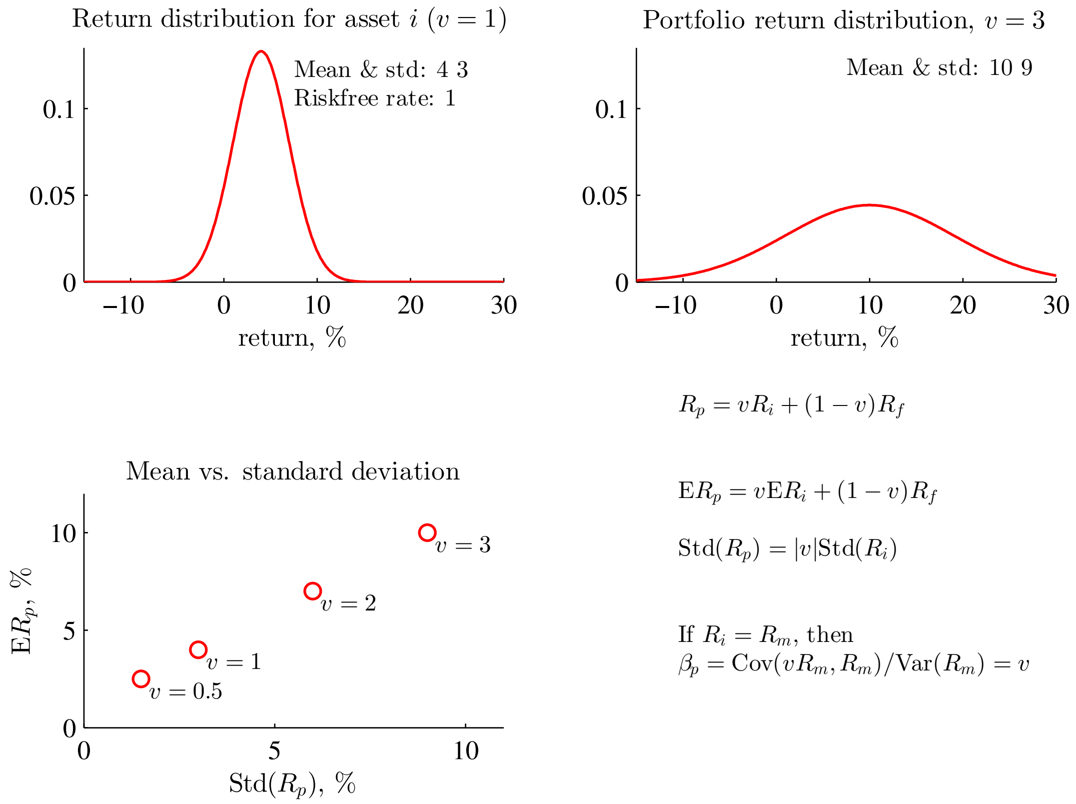


Figure 1.3: The effect of leverage on the portfolio return distribution

one riskfree asset (T-bill, say), the investor maximizes

$$\mathrm{E}\,U(R_p) = \mathrm{E}\,R_p - \frac{k}{2}\mathrm{Var}(R_p), \text{ where} \tag{1.21}$$

$$\begin{aligned} R_p &= vR_1 + (1 - v)R_f \\ &= vR_1^e + R_f. \end{aligned} \tag{1.22}$$

In the objective function k can be thought of as a measure of risk aversion.

Use the budget constraint in the objective function to get (using the fact that R_f is known)

$$\begin{aligned} \mathrm{E}\,U(R_p) &= \mathrm{E}(vR_1^e + R_f) - \frac{k}{2}\mathrm{Var}(vR_1^e + R_f) \\ &= v\mu_1^e + R_f - \frac{k}{2}v^2\sigma_{11}, \end{aligned} \tag{1.23}$$

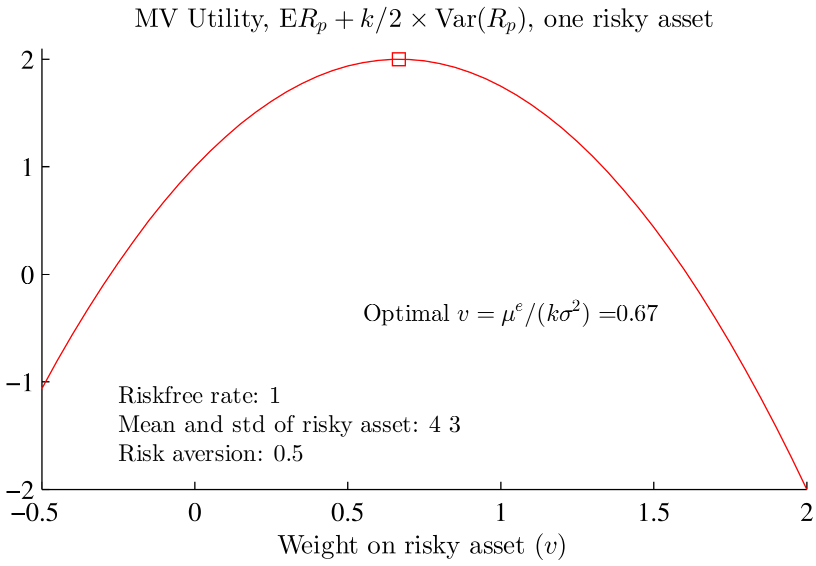


Figure 1.4: Portfolio choice

where σ_{11} denotes the variance of the risky asset.

The first order condition for an optimum is

$$0 = \partial \operatorname{E} U(R_p)/\partial v = \mu_1^e - kv\sigma_{11}, \tag{1.24}$$

so the optimal portfolio weight of the risky asset is

$$v = \frac{1}{k}\frac{\mu_1^e}{\sigma_{11}}. \tag{1.25}$$

The weight on the risky asset is increasing in the expected excess return of the risky asset, but decreasing in the risk aversion and variance.

Example 1.9 *(Portfolio choice) If $\mu_1^e = 3$, $\sigma_{11} = 9$ and $k = 0.5$, then $v \approx 0.67$. See Figure 1.4.*

This optimal solution implies that

$$\frac{\operatorname{E} R_p^e}{\operatorname{Var}(R_p)} = k, \tag{1.26}$$

where R_p is the portfolio return (1.22) obtained by using the optimal v (from (1.25)). It shows that an investor with a high risk aversion (k) will choose a portfolio with a high return compared to the volatility.

Proof. (of (1.26)) We have

$$\frac{\mathrm{E}\, R_p^e}{\mathrm{Var}(R_p)} = \frac{v\mu_1^e}{v^2\sigma_{11}} = \frac{\mu_1^e}{v\sigma_{11}},$$

which by using (1.25) gives (1.26). ■

1.4 Asset Classes

Table 1.2 shows the return ranking of some important subclasses of US equity and fixed income over the last decade. Figure 1.5

Much portfolio management is about trying to time these changes. The changes of the ranking—and in the returns—highlight both the opportunities (if you time it right) and risks (if you don't) with such an approach.

	6th		5th		4th		3rd		2nd		1st	
	6th		5th		4th		3rd		2nd		1st	
2003	TB	1.0	B	2.2	LV	21.1	LG	25.4	SG	73.7	SV	91.2
2004	TB	1.2	B	3.5	LG	5.5	SG	6.3	SV	22.5	LV	24.2
2005	SG	-3.1	LG	2.3	B	2.8	TB	3.0	SV	10.0	LV	12.8
2006	B	3.1	TB	4.8	SG	8.0	LG	8.0	LV	28.8	SV	30.6
2007	SV	-12.9	SG	-4.7	LV	3.8	TB	4.7	B	9.0	LG	11.7
2008	SG	-47.4	SV	-42.2	LG	-35.3	LV	-33.9	TB	1.7	B	13.7
2009	B	-3.6	TB	0.1	LV	23.1	SG	35.9	LG	36.8	SV	44.0
2010	TB	0.1	LV	5.8	B	5.9	LG	14.4	SG	25.5	SV	32.4
2011	SG	-20.3	LV	-14.2	SV	-8.3	TB	0.0	LG	3.2	B	9.8
2012	TB	0.1	B	2.0	SG	9.7	LG	15.8	SV	28.4	LV	34.6

Table 1.2: Ranking and return (in %) of asset classes, US. SG: small growth firms, SV: small value, LG: large growth, LV: large value, B: T-bonds, TB: T-bills.

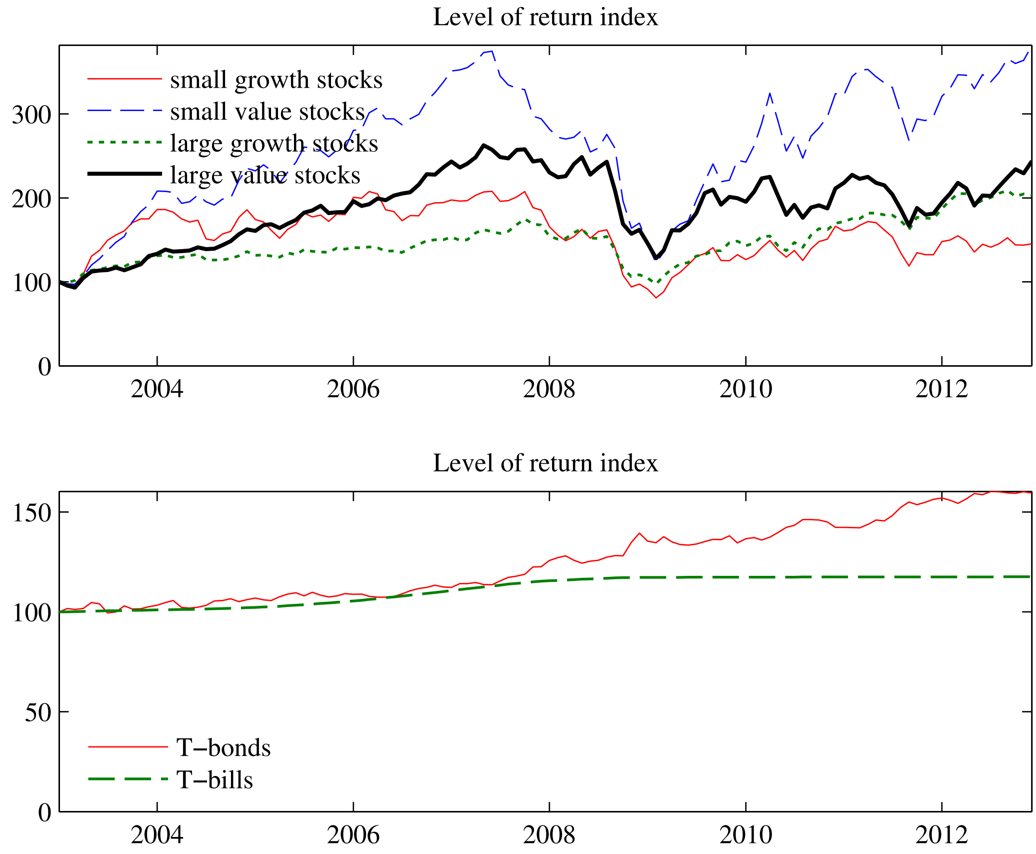


Figure 1.5: Performance of US equity and fixed income

A A Primer in Matrix Algebra

Let c be a scalar and define the matrices

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \text{ and } B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}.$$

Adding/subtracting a scalar to a matrix or multiplying a matrix by a scalar are both element by element

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} + c = \begin{bmatrix} A_{11}+c & A_{12}+c \\ A_{21}+c & A_{22}+c \end{bmatrix}$$
$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} c = \begin{bmatrix} A_{11}c & A_{12}c \\ A_{21}c & A_{22}c \end{bmatrix}.$$

Example A.1

$$\begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix} + 10 = \begin{bmatrix} 11 & 13 \\ 13 & 14 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix} 10 = \begin{bmatrix} 10 & 30 \\ 30 & 40 \end{bmatrix}.$$

Matrix *addition* (or subtraction) is element by element

$$A + B = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}+B_{11} & A_{12}+B_{12} \\ A_{21}+B_{21} & A_{22}+B_{22} \end{bmatrix}.$$

Example A.2 *(Matrix addition and subtraction)*

$$\begin{bmatrix} 10 \\ 11 \end{bmatrix} - \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 3 & -2 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 6 & 2 \end{bmatrix}$$

To turn a column into a row vector, use the *transpose* operator like in x'

$$x' = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}' = \begin{bmatrix} x_1 & x_2 \end{bmatrix}.$$

Similarly, transposing a matrix is like flipping it around the main diagonal

$$A' = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}' = \begin{bmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{bmatrix}.$$

Example A.3 *(Matrix transpose)*

$$\begin{bmatrix} 10 \\ 11 \end{bmatrix}' = \begin{bmatrix} 10 & 11 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}' = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

Matrix *multiplication* requires the two matrices to be conformable: the first matrix has as many columns as the second matrix has rows. Element ij of the result is the multiplication of the ith row of the first matrix with the jth column of the second matrix

$$AB = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}.$$

Multiplying a square matrix A with a column vector z gives a column vector

$$Az = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} A_{11}z_1 + A_{12}z_2 \\ A_{21}z_1 + A_{22}z_2 \end{bmatrix}.$$

Example A.4 *(Matrix multiplication)*

$$\begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & -2 \end{bmatrix} = \begin{bmatrix} 10 & -4 \\ 15 & -2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 17 \\ 26 \end{bmatrix}$$

For two column vectors x and z, the product $x'z$ is called the *inner product*

$$x'z = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = x_1z_1 + x_2z_2,$$

and xz' the *outer product*

$$xz' = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} z_1 & z_2 \end{bmatrix} = \begin{bmatrix} x_1z_1 & x_1z_2 \\ x_2z_1 & x_2z_2 \end{bmatrix}.$$

(Notice that xz does not work). If x is a column vector and A a square matrix, then the product $x'Ax$ is a *quadratic form*.

Example A.5 *(Inner product, outer product and quadratic form)*

$$\begin{bmatrix}10\\11\end{bmatrix}'\begin{bmatrix}2\\5\end{bmatrix}=\begin{bmatrix}10 & 11\end{bmatrix}\begin{bmatrix}2\\5\end{bmatrix}=75$$
$$\begin{bmatrix}10\\11\end{bmatrix}\begin{bmatrix}2\\5\end{bmatrix}'=\begin{bmatrix}10\\11\end{bmatrix}\begin{bmatrix}2 & 5\end{bmatrix}=\begin{bmatrix}20 & 50\\22 & 55\end{bmatrix}$$
$$\begin{bmatrix}10\\11\end{bmatrix}'\begin{bmatrix}1 & 3\\3 & 4\end{bmatrix}\begin{bmatrix}10\\11\end{bmatrix}=1244.$$

A matrix *inverse* is the closest we get to "dividing" by a matrix. The inverse of a matrix A, denoted A^{-1}, is such that

$$AA^{-1}=I \text{ and } A^{-1}A=I,$$

where I is the *identity matrix* (ones along the diagonal, and zeroes elsewhere). The matrix inverse is useful for solving systems of linear equations, $y=Ax$ as $x=A^{-1}y$.

Example A.6 *(Matrix inverse) We have*

$$\begin{bmatrix}-4/5 & 3/5\\3/5 & -1/5\end{bmatrix}\begin{bmatrix}1 & 3\\3 & 4\end{bmatrix}=\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}, \text{ so}$$
$$\begin{bmatrix}1 & 3\\3 & 4\end{bmatrix}^{-1}=\begin{bmatrix}-4/5 & 3/5\\3/5 & -1/5\end{bmatrix}.$$

Let z and x be $n\times 1$ vectors. The *derivative of the inner product* is $\partial(z'x)/\partial z=x$.

Example A.7 *(Derivative of an inner product) With $n=2$*

$$z'x=z_1x_1+z_2x_2, \text{ so } \frac{\partial(z'x)}{\partial z}=\frac{\partial(z_1x_1+z_2x_2)}{\begin{bmatrix}\partial z_1\\\partial z_2\end{bmatrix}}=\begin{bmatrix}x_1\\x_2\end{bmatrix}.$$

Let x be $n\times 1$ and A a symmetric $n\times n$ matrix. The *derivative of the quadratic form* is $\partial(x'Ax)/\partial x=2Ax$.

Example A.8 *(Derivative of a quadratic form) With $n=2$, the quadratic form is*

$$x'Ax=\begin{bmatrix}x_1 & x_2\end{bmatrix}\begin{bmatrix}A_{11} & A_{12}\\A_{12} & A_{22}\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=x_1^2A_{11}+x_2^2A_{22}+2x_1x_2A_{12}.$$

The derivatives with respect to x_1 and x_2 are

$$\frac{\partial(x'Ax)}{\partial x_1} = 2x_1A_{11} + 2x_2A_{12} \text{ and } \frac{\partial(x'Ax)}{\partial x_2} = 2x_2A_{22} + 2x_1A_{12}, \text{ or}$$

$$\frac{\partial(x'Ax)}{\begin{bmatrix}\partial x_1 \\ \partial x_2\end{bmatrix}} = 2\begin{bmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

B A Primer in Optimization

You want to choose x and y to minimize

$$L = (x-2)^2 + (4y+3)^2,$$

then we have to find the values of x and y that satisfy the *first order conditions* $\partial L/\partial x = \partial L/\partial y = 0$. These conditions are

$$\begin{aligned} 0 &= \partial L/\partial x = 2(x-2) \\ 0 &= \partial L/\partial y = 8(4y+3), \end{aligned}$$

which clearly requires $x = 2$ and $y = -3/4$. In this particular case, the first order condition with respect to x does not depend on y, but that is not a general property. In this case, this is the unique solution—but in more complicated problems, the first order conditions could be satisfied at different values of x and y.

See Figure B.1 for an illustration.

If you want to add a restriction to the minimization problem, say

$$x + 2y = 3,$$

then we can proceed in two ways. The first is to simply substitute for $x = 3 - 2y$ in L to get

$$L = (1-2y)^2 + (4y+3)^2,$$

with first order condition

$$0 = \partial L/\partial y = -4(1-2y) + 8(4y+3) = 40y + 20,$$

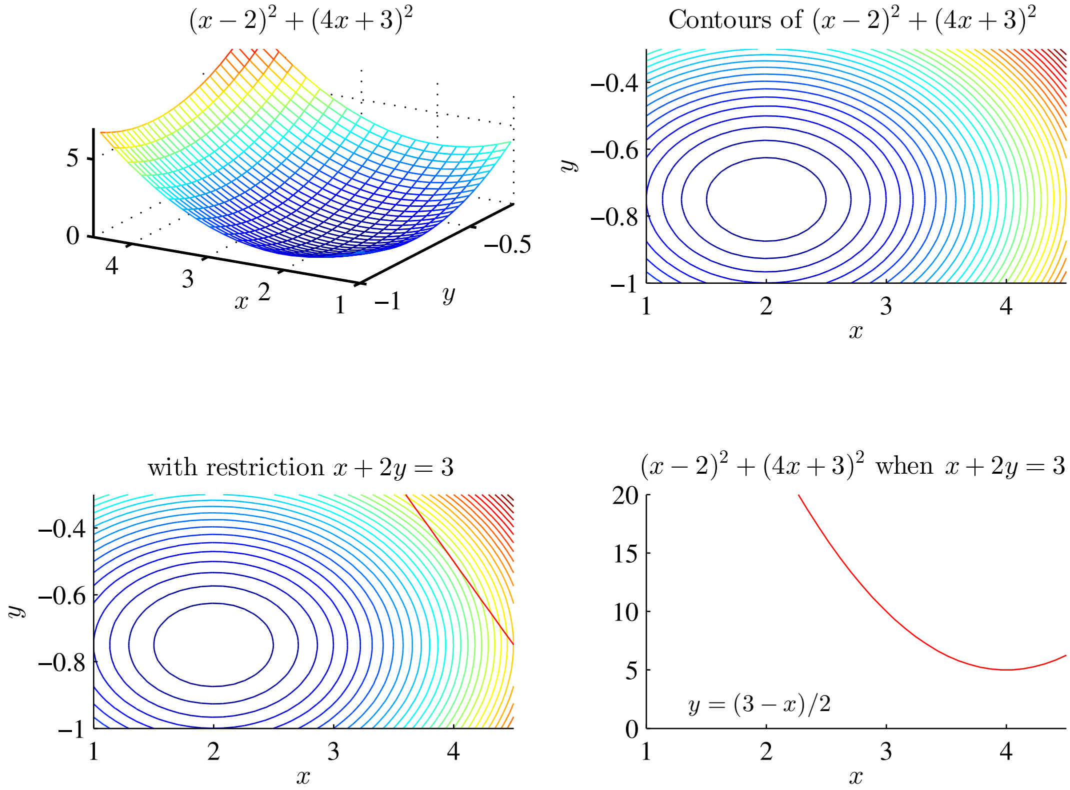


Figure B.1: Minimization problem

which requires $y = -1/2$. (We could equally well have substituted for y). This is also the unique solution.

The second method is to use a *Lagrangian*. The problem is then to choose x, y, *and* λ to minimize

$$L = (x-2)^2 + (4y+3)^2 + \lambda\,(3-x-2y)\,.$$

The term multiplying λ is the restriction. The first order conditions are now

$$\begin{aligned}
0 &= \partial L/\partial x = 2(x-2) - \lambda \\
0 &= \partial L/\partial y = 8(4y+3) - 2\lambda \\
0 &= \partial L/\partial \lambda = 3 - x - 2y.
\end{aligned}$$

The first two conditions say

$$x = \lambda/2 + 2$$
$$y = \lambda/16 - 3/4,$$

so we need to find λ. To do that, use these latest expressions for x and y in the third first order condition (to substitute for x and y)

$$3 = \lambda/2 + 2 + 2\,(\lambda/16 - 3/4) = \lambda 5/8 + 1/2, \text{ so}$$
$$\lambda = 4.$$

Finally, use this to calculate x and y as

$$x = 4 \text{ and } y = -1/2.$$

Notice that this is the same solution as before ($y = -1/2$) and that the restriction holds ($4 + 2(-1/2) = 3$). This second method is clearly a lot clumsier in my example, but it pays off when the restriction(s) become complicated.

Bibliography

Elton, E. J., M. J. Gruber, S. J. Brown, and W. N. Goetzmann, 2010, *Modern portfolio theory and investment analysis*, John Wiley and Sons, 8th edn.

Fabozzi, F. J., S. M. Focardi, and P. N. Kolm, 2006, *Financial modeling of the equity market*, Wiley Finance.

2 Mean-Variance Frontier

Reference: Elton, Gruber, Brown, and Goetzmann (2010) 4–6; Fabozzi, Focardi, and Kolm (2006) 4

2.1 Mean-Variance Frontier of Risky Assets

To calculate a point on the mean-variance frontier, we have to find the portfolio that minimizes the portfolio variance, $\mathrm{Var}(R_p)$, for a given expected return, μ^*. The problem is thus

$$\begin{aligned} &\min_{w_i} \mathrm{Var}(R_p) \text{ subject to} \\ &\mathrm{E}\, R_p = \textstyle\sum_{i=1}^n w_i \mu_i = \mu^* \text{ and } \sum_{i=1}^n w_i = 1. \end{aligned} \tag{2.1}$$

Let Σ be the covariance matrix of the asset returns. The portfolio variance is then calculated as

$$\mathrm{Var}(R_p) = \mathrm{Var}(\textstyle\sum_{i=1}^n w_i R_i) = w' \Sigma w. \tag{2.2}$$

The whole mean-variance frontier is generated by solving this problem for different values of the expected return (μ^*). The results are typically shown in a figure with the standard deviation on the horizontal axis and the required return on the vertical axis. The *efficient frontier* is the upper leg of the curve. Reasonably, a portfolio on the lower leg is dominated by one on the upper leg at the same volatility (since it has a higher expected return). See Figure 2.1 for an example.

Remark 2.1 *(Only two assets) In the (empirically uninteresting) case of only two assets, the MV frontier can be calculated by simply calculating the mean and variance*

$$\begin{aligned} \mathrm{E}\, R_p &= w\mu_1 + (1-w)\mu_2 \\ \mathrm{Var}(R_p) &= w\sigma_{11} + (1-w)^2\sigma_{22} + 2w(1-w)\sigma_{12}. \end{aligned}$$

at a set of different portfolio weights (for instance, $w = (0, 0.25, 0.5, 0.75, 1)$.*) The reason is that, with only two assets, both assets are on the MV frontier—so no explicit*

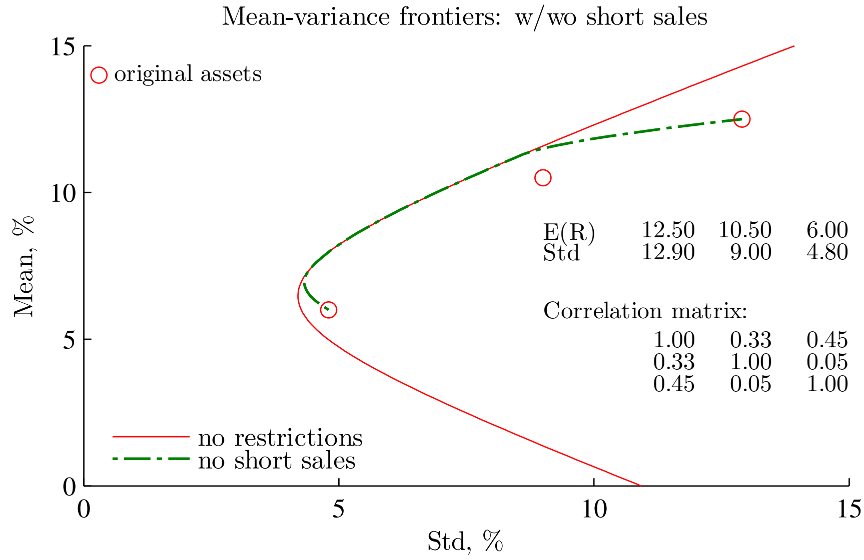


Figure 2.1: Mean-variance frontiers

minimization is needed. See Figures 2.2–2.3 for examples.

It is (relatively) straightforward to calculate the mean-variance frontier if there are no other constraints: it just takes some linear algebra—see Section 2.1.2. See Figure 2.5 for an example.

There are sometimes *additional restrictions*, for instance,

$$\text{no short sales: } w_i \geq 0. \tag{2.3}$$

We then have to apply some explicit numerical minimization algorithm to find portfolio weights. Algorithms that solve quadratic problems are best suited (this is a quadratic problem—see (2.2)). See Figure 2.1 for an example. Other commonly used restrictions are that the new weights should not deviate too much from the old (when rebalancing)—in an effort to reduce trading costs

$$|w_i^{new} - w_i^{old}| < U_i, \tag{2.4}$$

or that the portfolio weights must be between some boundaries

$$L_i \leq w_i \leq U_i. \tag{2.5}$$

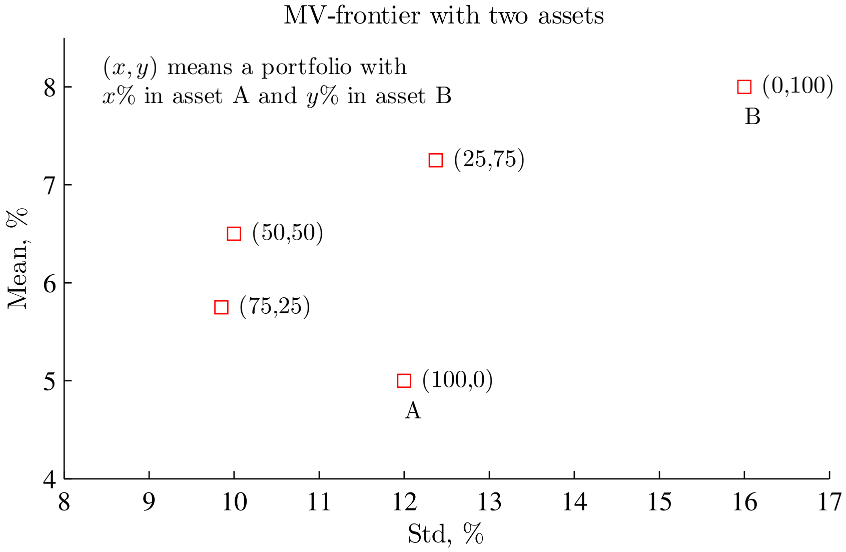


Figure 2.2: Mean-variance frontiers for two risky assets.

Consider what happens when we *add assets to the investment opportunity set*. The old mean-variance frontier is, of course, still obtainable: we can always put zero weights on the new assets. In most cases, we can do better than that so the mean-variance frontier is moved to the left (lower volatility at the same expected return). See Figure 2.4 for an example.

2.1.1 The Shape of the MV Frontier of Risky Assets

This section discusses how the shape of the MV frontier depends on the correlation of the assets. For simplicity, only two assets are used but the general findings hold also when there are more assets.

With intermediate correlations ($-1 < \rho < 1$) the mean-variance frontier is a hyperbola—see Figure 2.6. Notice that the mean–volatility trade-off improves as the correlation decreases: a lower correlation means that we get a lower portfolio standard deviation at the same expected return—at least for the efficient frontier (above the bend).

When the assets are *perfectly correlated* ($\rho = 1$), then the frontier is a pair of two straight lines—see Figures 2.7–2.8. The efficient frontier is clearly the upper leg. However, if short sales are ruled out then the MV frontier is just a straight line connecting the two assets. The intuition is that a perfect correlation means that the second asset is a linear

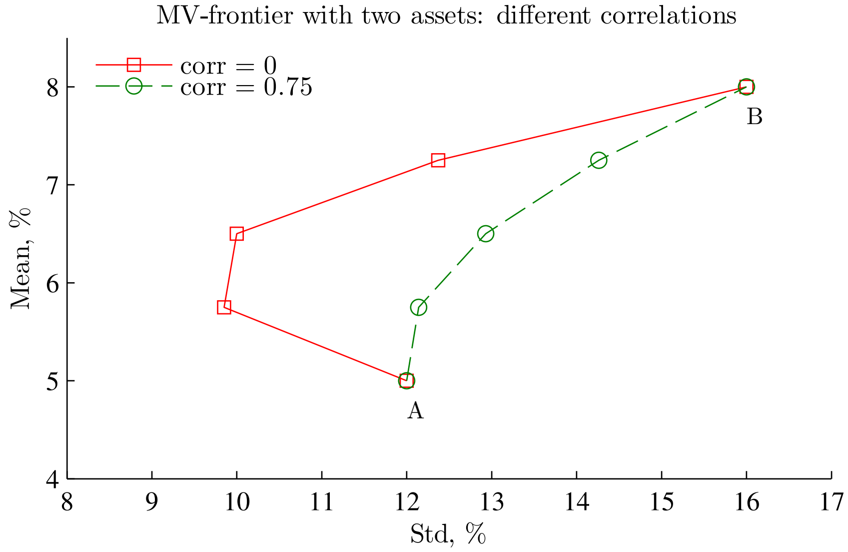


Figure 2.3: Mean-variance frontiers for two risky assets, different correlations.

transformation of the first ($R_2 = a + bR_1$), so changing the portfolio weights essentially means forming just another linear combination of the first asset. In particular, there are no diversification benefits. In fact, the case of a perfect (positive) correlation is a limiting case: a combination of two assets can never have higher standard deviation than the line connecting them in the $\sigma \times \mathrm{E}\, R$ space.

Also when the assets are *perfectly negatively correlated* ($\rho = -1$), then the MV frontier is a pair of straight lines, see Figures 2.7–2.8. In contrast to the case with a perfect positive correlation, this is true also when short sales are ruled out. This means, for instance, that we can combine the two assets (with positive weights) to get a riskfree portfolio.

Proof. (of the MV shapes with 2 assets*) With a perfect correlation ($\rho = 1$) the standard deviation can be rearranged. Suppose the portfolio weights are positive (no short sales). Then we get

$$\begin{aligned}\sigma_p &= \left[w_1^2\sigma_{11} + (1-w_1)^2\sigma_{22} + 2w_1(1-w_1)\sigma_1\sigma_2\right]^{1/2}\\ &= \left\{[w_1\sigma_1 + (1-w_1)\sigma_2]^2\right\}^{1/2}\\ &= w_1\sigma_1 + (1-w_1)\sigma_2.\end{aligned}$$

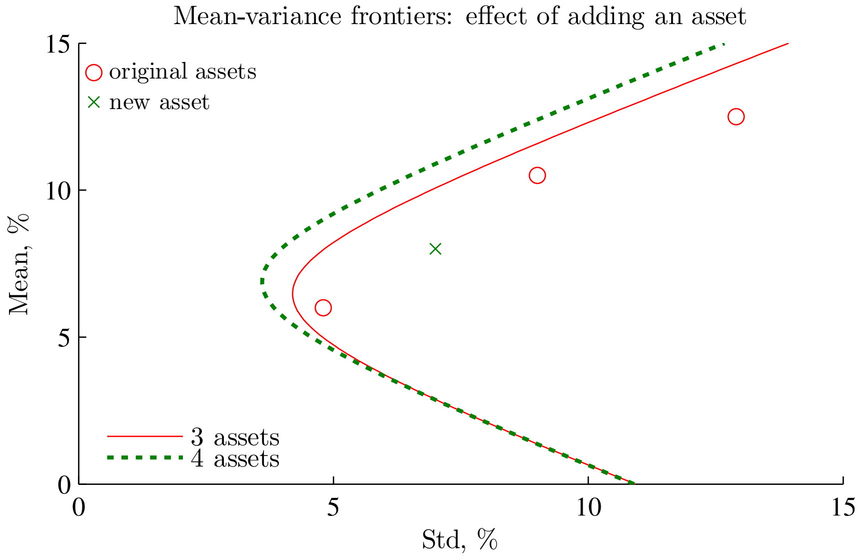


Figure 2.4: Mean-variance frontiers

We can rearrange this expression as $w_1 = \left(\sigma_p - \sigma_2\right) / \left(\sigma_1 - \sigma_2\right)$ which we can use in the expression for the expected return to get

$$\mathrm{E}\, R_p = \frac{\sigma_p - \sigma_2}{\sigma_1 - \sigma_2} \left(\mathrm{E}\, R_1 - \mathrm{E}\, R_2\right) + \mathrm{E}\, R_2.$$

This shows that the mean-variance frontier is just a straight line (if there are no short sales). We get a riskfree portfolio ($\sigma_p = 0$) if $w_1 = \sigma_2 / \left(\sigma_2 - \sigma_1\right)$.

With a perfectly negative correlation ($\rho = -1$) the standard deviation can be rearranged as follows (assuming positive weights)

$$\begin{aligned}
\sigma_p &= \left[w_1^2 \sigma_{11} + \left(1 - w_1\right)^2 \sigma_{22} - 2 w_1 \left(1 - w_1\right) \sigma_1 \sigma_2\right]^{1/2} \\
&= \begin{cases} \left\{\left[w_1 \sigma_1 - \left(1 - w_1\right) \sigma_2\right]^2\right\}^{1/2} = w_1 \sigma_1 - \left(1 - w_1\right) \sigma_2 \text{ if } [] \geq 0 \\ \left\{\left[-w_1 \sigma_1 + \left(1 - w_1\right) \sigma_2\right]^2\right\}^{1/2} = -w_1 \sigma_1 + \left(1 - w_1\right) \sigma_2 \text{ if } [] \geq 0. \end{cases}
\end{aligned}$$

The 2nd expression is -1 times the 1st expression. Only one can be positive at each time. Both have same form as in case with $\rho = 1$, so both generate linear relation: $\mathrm{E}\left(R_p\right) = a + b\sigma_p$—but with different slopes. We get a riskfree portfolio ($\sigma_p = 0$) if $w_1 = \sigma_2 / (\sigma_1 + \sigma_2)$. ■

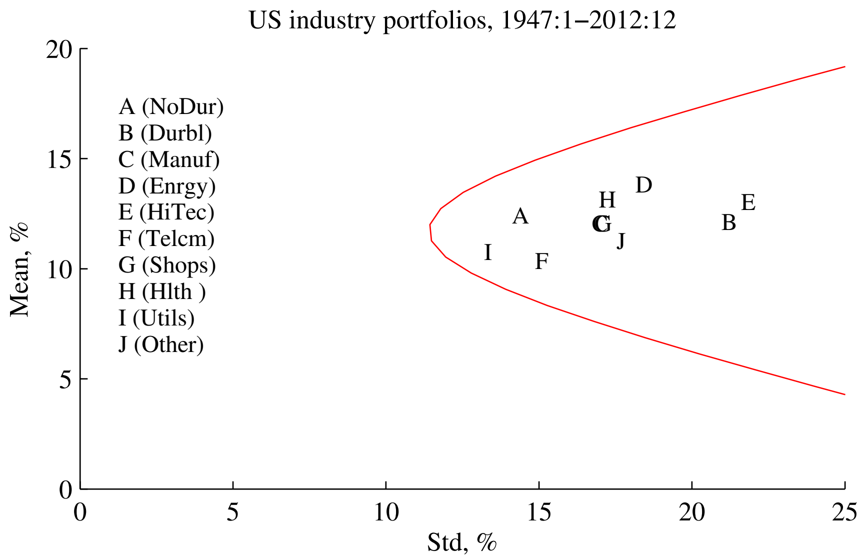


Figure 2.5: M-V frontier from US industry indices

2.1.2 Calculating the MV Frontier of Risky Assets: No Restrictions

When there are no restrictions on the portfolio weights, then there are two ways of finding a point on the mean-variance frontier: let a numerical optimization routine do the work or use some simple matrix algebra. The section demonstrates the second approach.

To simplify the following equations, define the scalars A, B and C as

$$A = \mu' \Sigma^{-1} \mu,\ B = \mu' \Sigma^{-1} \mathbf{1}, \text{ and } C = \mathbf{1}' \Sigma^{-1} \mathbf{1}, \tag{2.6}$$

where $\mathbf{1}$ is a (column) vector of ones and μ' is the transpose of the column vector μ. Then, calculate the scalars (for a given required return μ^*)

$$\lambda = \frac{C\mu^* - B}{AC - B^2} \text{ and } \delta = \frac{A - B\mu^*}{AC - B^2}. \tag{2.7}$$

The weights for a portfolio on the MV frontier of risky assets (at a given required return μ^*) are then

$$w = \Sigma^{-1}(\mu\lambda + \mathbf{1}\delta). \tag{2.8}$$

Using this in (2.2) gives the variance (take the square root to get the standard deviation). We can trace out the entire MV frontier, by repeating this calculations for different values

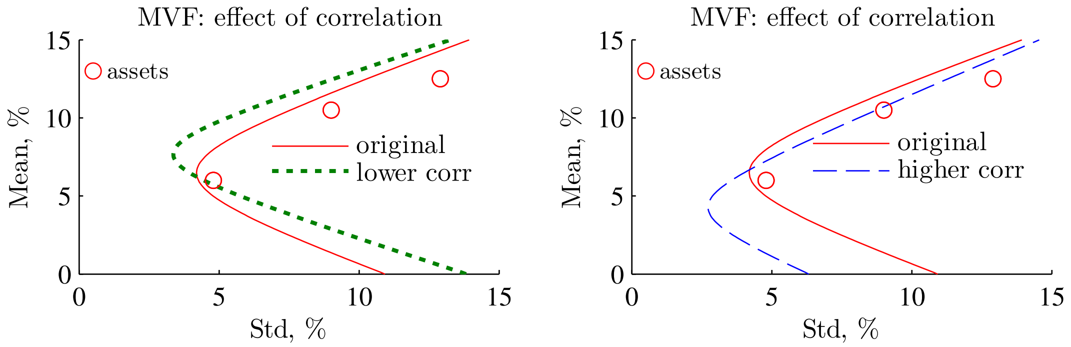


Figure 2.6: Mean-variance frontiers for normal and high correlations.

of the required return and then connecting the dots. In the std×mean space, the efficient frontier (the upper part) is concave. See Figure 2.1 for an example.

Example 2.2 *(Transpose of a matrix) Consider the following examples*

$$\begin{bmatrix}1\\3\\5\end{bmatrix}' = \begin{bmatrix}1 & 3 & 5\end{bmatrix}, \begin{bmatrix}1 & 2\\3 & 4\\5 & 6\end{bmatrix}' = \begin{bmatrix}1 & 3 & 5\\2 & 4 & 6\end{bmatrix} \text{ and } \begin{bmatrix}1 & 2\\2 & 4\end{bmatrix}' = \begin{bmatrix}1 & 2\\2 & 4\end{bmatrix}.$$

Transposing a symmetric matrix does nothing, that is, if A is symmetric, then $A' = A$.

Proof. (of (2.6)–(2.8)) We set up this as a Lagrangian problem

$$L = (w_1^2\sigma_{11} + w_2^2\sigma_{22} + 2w_1w_2\sigma_{12})/2 + \lambda(\mu^* - w_1\mu_1 - w_2\mu_2) + \delta(1 - w_1 - w_2).$$

The first order condition with respect to w_i is $\partial L/\partial w_i = 0$, that is,

$$\begin{aligned}
\text{for } w_1 &: w_1\sigma_{11} + w_2\sigma_{12} - \lambda\mu_1 - \delta = 0,\\
\text{for } w_2 &: w_1\sigma_{12} + w_2\sigma_{22} - \lambda\mu_2 - \delta = 0.
\end{aligned}$$

In matrix notation these first order conditions are

$$\begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{bmatrix}\begin{bmatrix}w_1\\w_2\end{bmatrix} - \lambda\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} - \delta\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}.$$

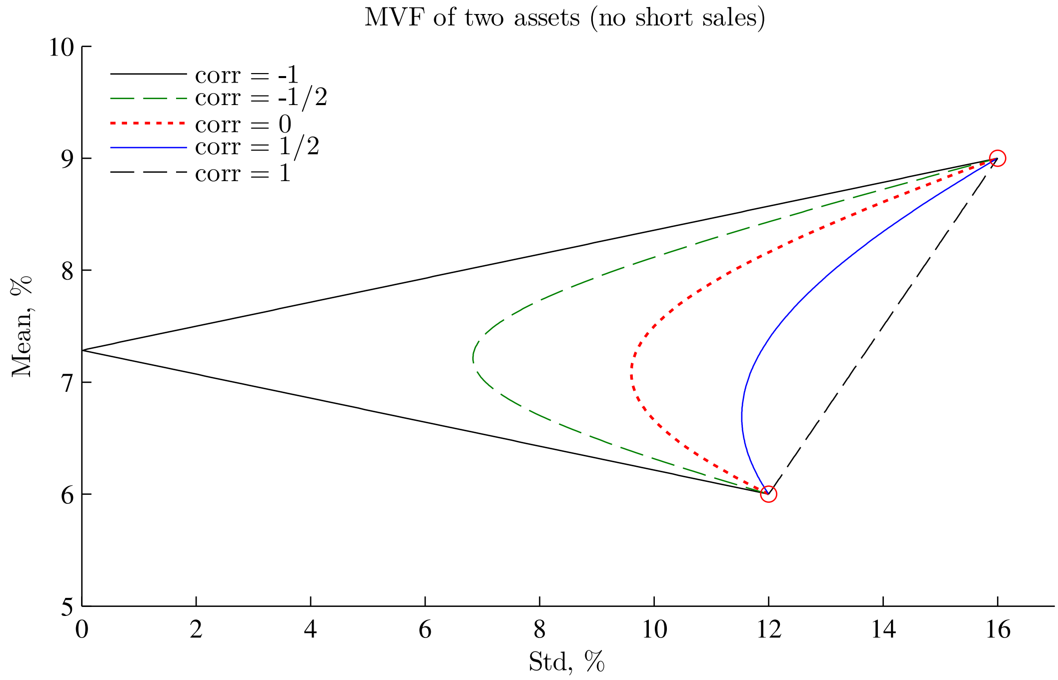


Figure 2.7: Mean-variance frontiers for two risky assets when short sales are not allowed: different correlations. The two assets are indicated by circles.

We can solve these equations for w_1 and w_2 as

$$\begin{aligned}
\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} &= \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix} \left(\lambda \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \delta \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) \\
&= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}^{-1} \left(\lambda \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \delta \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) \\
w &= \Sigma^{-1}(\lambda\mu + \delta\mathbf{1}),
\end{aligned}$$

where $\mathbf{1}$ is a column vector of ones. The first order conditions for the Lagrange multipliers are (of course)

$$\begin{aligned}
\text{for } \lambda &: \mu^* - w_1\mu_1 - w_2\mu_2 = 0, \\
\text{for } \delta &: 1 - w_1 - w_2 = 0.
\end{aligned}$$

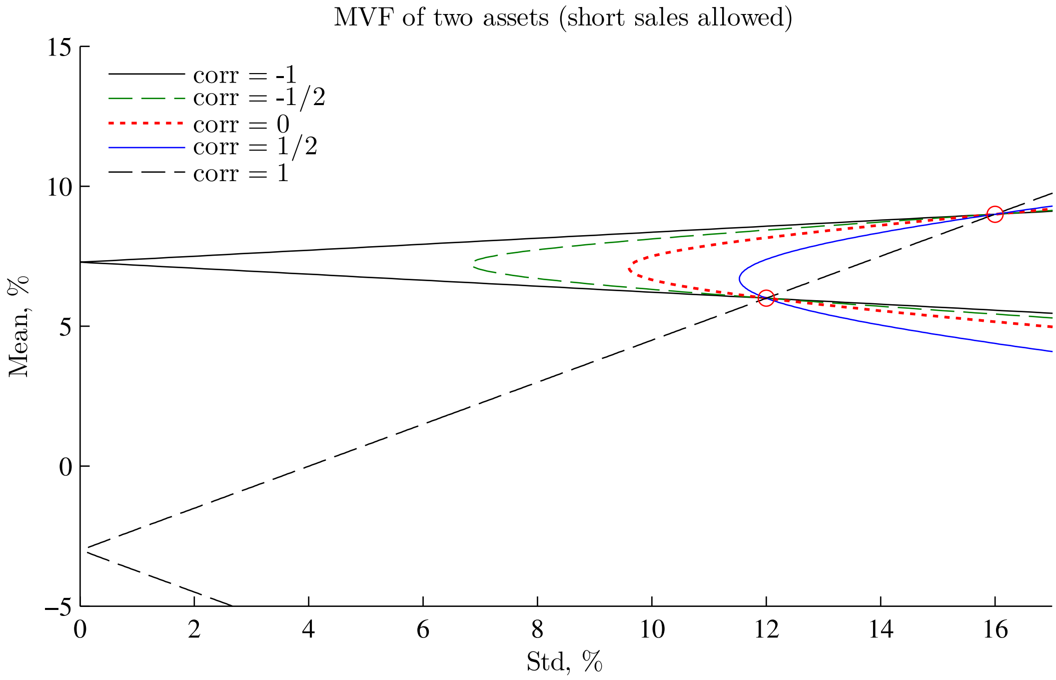


Figure 2.8: Mean-variance frontiers for two risky assets: different correlations. The two assets are indicated by circles. Points between the two assets can be generated with positive portfolio weights (no short sales).

In matrix notation, these conditions are

$$\mu^* = \mu' w \text{ and } 1 = \mathbf{1}' w.$$

Stack these into a 2×1 vector and substitute for w

$$\begin{aligned}
\begin{bmatrix} \mu^* \\ 1 \end{bmatrix} &= \begin{bmatrix} \mu' \\ \mathbf{1}' \end{bmatrix} w \\
&= \begin{bmatrix} \mu' \\ \mathbf{1}' \end{bmatrix} \Sigma^{-1}(\lambda\mu + \delta\mathbf{1}) \\
&= \begin{bmatrix} \mu'\Sigma^{-1}\mu & \mu'\Sigma^{-1}\mathbf{1} \\ \mathbf{1}'\Sigma^{-1}\mu & \mathbf{1}'\Sigma^{-1}\mathbf{1} \end{bmatrix} \begin{bmatrix} \lambda \\ \delta \end{bmatrix} \\
&= \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} \lambda \\ \delta \end{bmatrix}.
\end{aligned}$$

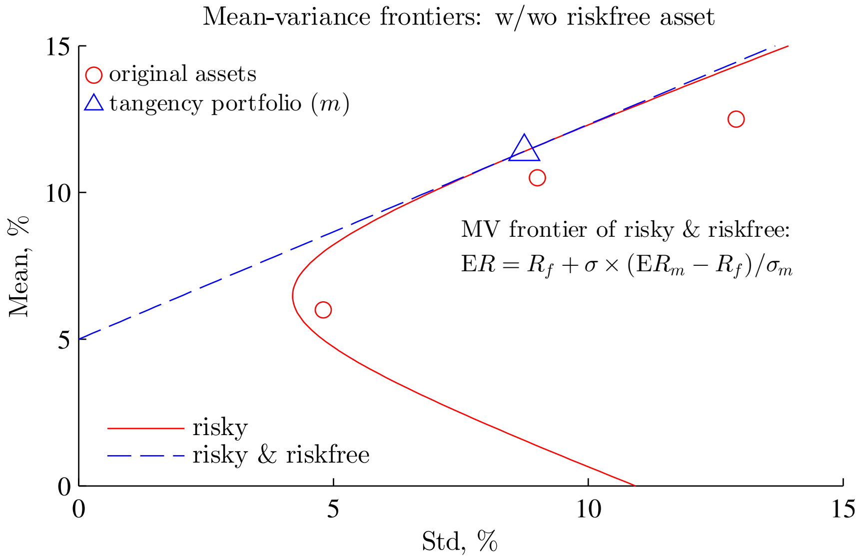


Figure 2.9: Mean-variance frontiers

Solve for λ and δ as

$$\lambda = \frac{C\mu^* - B}{AC - B^2} \text{ and } \delta = \frac{A - B\mu^*}{AC - B^2}.$$

Use this in the expression for w above. ■

2.2 Mean-Variance Frontier of Riskfree and Risky Assets

We now add a riskfree asset with return R_f. With two risky assets, the portfolio return is

$$\begin{aligned} R_p &= w_1 R_1 + w_2 R_2 + (1 - w_1 - w_2) R_f \\ &= w_1 (R_1 - R_f) + w_2 (R_2 - R_f) + R_f \\ &= w_1 R_1^e + w_2 R_2^e + R_f, \end{aligned} \tag{2.9}$$

where R_i^e is the excess return of asset i. We denote the corresponding expected excess return by μ_i^e (so $\mu_i^e = \mathrm{E}\, R_i^e$).

The minimization problem is now

$$\min_{w_1,w_2} (w_1^2 \sigma_{11} + w_2^2 \sigma_{22} + 2 w_1 w_2 \sigma_{12})/2 \text{ subject to} \tag{2.10}$$

$$w_1 \mu_1^e + w_2 \mu_2^e + R_f = \mu^*.$$

Notice that we don't need any restrictions on the sum of weights: the investment in the riskfree rate automatically makes the overall sum equal to unity.

With more assets, the minimization problem is

$$\begin{aligned} &\min_{w_i} \mathrm{Var}(R_p) \text{ subject to} \\ &\mathrm{E}\, R_p = \textstyle\sum_{i=1}^{n} w_i \mu_i^e + R_f = \mu^*, \end{aligned} \tag{2.11}$$

where the portfolio variance is calculated as usual

$$\mathrm{Var}(R_p) = \mathrm{Var}(\textstyle\sum_{i=1}^{n} w_i R_i) = w' \Sigma w. \tag{2.12}$$

When there are no additional constraints, then we can find an explicit solution in terms of some matrices and vectors—see Section 2.2.1. In all other cases, we need to apply an explicit numerical minimization algorithm (preferably for quadratic models).

2.2.1 Calculating the MV Frontier of Riskfree and Risky Assets: No Restrictions

The weights (of the risky assets) for a portfolio on the MV frontier (at a given required return μ^*) are

$$w = \frac{\mu^* - R_f}{(\mu^e)' \Sigma^{-1} \mu^e} \Sigma^{-1} \mu^e, \tag{2.13}$$

where R_f is the riskfree rate and μ^e the vector of mean excess returns ($\mu - R_f$). The weight on the riskfree asset is $1 - \mathbf{1}'w$.

Using this in (2.2) gives the variance (take the square root to get the standard deviation). We can trace out the entire MV frontier, by repeating this calculations for different values of the required return and then connecting the dots. In the std×mean space, the efficient frontier (the upper part) is just a line. See Figure 2.9 for an example.

Proof. (of (2.13)) Define the Lagrangian problem

$$L = (w_1^2 \sigma_{11} + w_2^2 \sigma_{22} + 2 w_1 w_2 \sigma_{12})/2 + \lambda(\mu^* - w_1 \mu_1^e - w_2 \mu_2^e - R_f).$$

The first order condition with respect to w_i is $\partial L / \partial w_i = 0$, so

$$\begin{aligned} &\text{for } w_1 : w_1 \sigma_{11} + w_2 \sigma_{12} - \lambda \mu_1^e = 0, \\ &\text{for } w_2 : w_1 \sigma_{12} + w_2 \sigma_{22} - \lambda \mu_2^e = 0. \end{aligned}$$

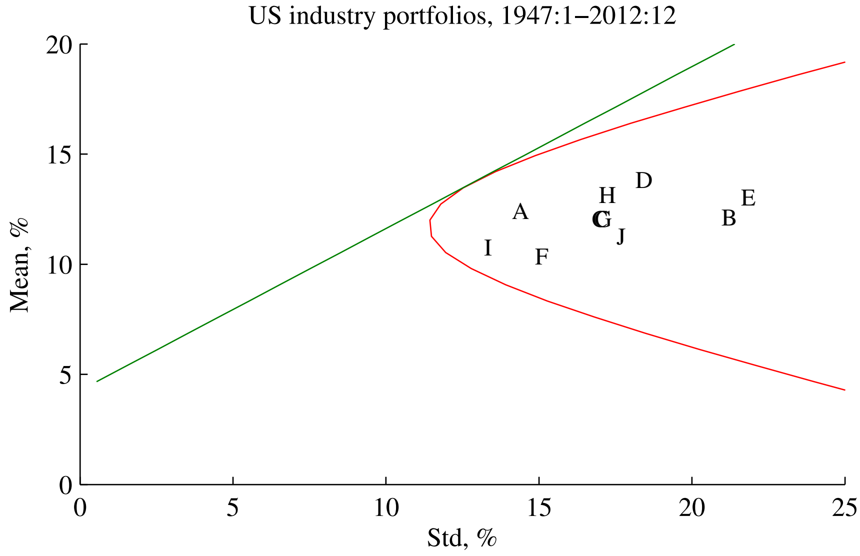


Figure 2.10: M-V frontier from US industry indices

It is then immediate that we can write them in matrix form as

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} - \lambda \begin{bmatrix} \mu_1^e \\ \mu_2^e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \text{ so}$$

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}^{-1} \lambda \begin{bmatrix} \mu_1^e \\ \mu_2^e \end{bmatrix}, \text{ or}$$

$$w = \Sigma^{-1} \lambda \mu^e.$$

The first order condition for the Lagrange multiplier is (in matrix form)

$$\mu^* = w' \mu^e + R_f.$$

Combine to get

$$\mu^* = \lambda (\mu^e)' \Sigma^{-1} \mu^e + R_f, \text{ so}$$

$$\lambda = \frac{\mu^* - R_f}{(\mu^e)' \Sigma^{-1} \mu^e}.$$

Use in the above expression for w. ■

2.2.2 Tangency Portfolio

The MV frontier for risky assets and the frontier for risky+riskfree assets are tangent at one point—called the *tangency portfolio*. In this case the portfolio weights (2.8) and (2.13) coincide. Therefore, the portfolio weights (2.13) must sum to unity (so the weight on the riskfree asset is zero) at this value of the required return, μ^*. This helps use to understand what the expected excess return on the tangency portfolio is—which if used in (2.13) gives the portfolio weights of the tangency portfolio

$$w = \frac{\Sigma^{-1}\mu^e}{\mathbf{1}'\Sigma^{-1}\mu^e}. \tag{2.14}$$

Proof. (of (2.14)) Put the sum of the portfolio weights in (2.13) equal to one

$$\mathbf{1}'w = \frac{\mu^* - R_f}{(\mu^e)'\Sigma^{-1}\mu^e}\mathbf{1}'\Sigma^{-1}\mu^e = 1,$$

which only happens if

$$\mu^* - R_f = \frac{(\mu^e)'\Sigma^{-1}\mu^e}{\mathbf{1}'\Sigma^{-1}\mu^e}.$$

Using in (2.13) gives (2.14). ■

2.3 Examples of Portfolio Weights from MV Calculations

With 2 risky assets and 1 riskfree asset the portfolio weights satisfy (2.13). We can write this as

$$w = \lambda \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22}\mu_1^e - \sigma_{12}\mu_2^e \\ \sigma_{11}\mu_2^e - \sigma_{12}\mu_1^e \end{bmatrix}, \tag{2.15}$$

where $\lambda > 0$ if we limit our attention to the efficient part where $\mu^* > R_f$. (This follows from the fact that $(\mu^e)'\Sigma^{-1}\mu^e > 0$ since Σ^{-1} is positive definite, because Σ is). We can then discuss some general properties of all portfolios in the efficient set.

Simple Case 1: Uncorrelated Assets ($\sigma_{12} = 0$)

From (2.15) we then get

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \lambda \begin{bmatrix} \mu_1^e/\sigma_{11} \\ \mu_2^e/\sigma_{22} \end{bmatrix}. \tag{2.16}$$

Suppose that $\lambda > 0$ (efficient part of the MV frontier) and that both excess returns are positive. In that case we have the following.

First, both weights are positive. The intuition is that uncorrelated assets make it efficient to diversify (to get the same expected return, but at a lower variance).

Second, the asset with the highest μ_i^e/σ_{ii} ratio has the highest portfolio weight. The intuition is that an asset with a high excess return and/or low volatility is an efficient way to achieve a low volatility at a given mean return.

Notice that increasing μ_i^e/σ_{ii} does not guarantee that the actual weight on asset i increases (because λ changes too). For instance, an increase in the expected return of an asset may allow us to shift assets towards the riskfree asset (and still get the same expected portfolio return, but lower variance).

Example 2.3 *(Portfolio weights with uncorrelated assets) When* $(\mu_1^e, \mu_2^e) = (0.07, 0.07)$*, the correlation is zero,* $(\sigma_{11}, \sigma_{22}) = (1, 1)$*, and* $\mu^* - R = 0.09$*, then (2.16) gives*

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 9.18 \begin{bmatrix} 0.07 \\ 0.07 \end{bmatrix} = \begin{bmatrix} 0.64 \\ 0.64 \end{bmatrix}.$$

If we change to $(\mu_1^e, \mu_2^e) = (0.09, 0.07)$*, then*

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 6.92 \begin{bmatrix} 0.09 \\ 0.07 \end{bmatrix} = \begin{bmatrix} 0.62 \\ 0.48 \end{bmatrix}.$$

If we instead change to $(\sigma_{11}, \sigma_{22}) = (1/2, 1)$*, then*

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 6.12 \begin{bmatrix} 0.14 \\ 0.07 \end{bmatrix} = \begin{bmatrix} 0.86 \\ 0.43 \end{bmatrix}.$$

Simple Case 2: Same Variances (but Correlation)

Let $\sigma_{11} = \sigma_{22} = 1$ (as a normalization), so the covariance becomes the correlation $\sigma_{12} = \rho$ where $-1 < \rho < 1$.

From (2.15) we then get

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \lambda \frac{1}{1-\rho^2} \begin{bmatrix} \mu_1^e - \rho\mu_2^e \\ \mu_2^e - \rho\mu_1^e \end{bmatrix}. \tag{2.17}$$

Suppose that $\lambda > 0$ (efficient part of the MV frontier) and that both excess returns are positive. In that case, we have the following.

First, both weights are positive if the returns are negatively correlated ($\rho < 0$). The intuition is that a negative correlation means that the assets "hedge" each other (even better than diversification), so the investor would like to hold both of them to reduce the overall risk.

Second, if $\rho > 0$ and μ_1^e is considerably higher than μ_2^e (so $\mu_2^e < \rho\mu_1^e$, which also implies $\mu_1^e > \rho\mu_2^e$), then $w_1 > 0$ but $w_2 < 0$. The intuition is that a positive correlation reduces the gain from holding both assets (they don't hedge each other, and there is relatively little diversification to be gained if the correlation is high). On top of this, asset 1 gives a higher expected return, so it is optimal to sell asset 2 short (essentially a risky "loan" which allows the investor to buy more of asset 1).

Example 2.4 *(Portfolio weights with correlated assets) When* $(\mu_1^e, \mu_2^e) = (0.07, 0.07)$*,* $\rho = 0.8$*, and* $\mu^* - R = 0.09$*, then (2.16) gives*

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 16.53 \begin{bmatrix} 0.039 \\ 0.039 \end{bmatrix} = \begin{bmatrix} 0.64 \\ 0.64 \end{bmatrix}.$$

This is the same as in the previous example. If we change to $(\mu_1^e, \mu_2^e) = (0.09, 0.07)$*, then we get*

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 11.10 \begin{bmatrix} 0.094 \\ -0.006 \end{bmatrix} = \begin{bmatrix} 1.05 \\ -0.06 \end{bmatrix}.$$

If we also change to $\rho = -0.8$*, then we get*

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 1.40 \begin{bmatrix} 0.406 \\ 0.394 \end{bmatrix} = \begin{bmatrix} 0.57 \\ 0.55 \end{bmatrix}.$$

These two last solutions are very different from the previous example.

Bibliography

Elton, E. J., M. J. Gruber, S. J. Brown, and W. N. Goetzmann, 2010, *Modern portfolio theory and investment analysis*, John Wiley and Sons, 8th edn.

Fabozzi, F. J., S. M. Focardi, and P. N. Kolm, 2006, *Financial modeling of the equity market*, Wiley Finance.

3 Index Models

Reference: Elton, Gruber, Brown, and Goetzmann (2010) 7–8, 11

3.1 The Inputs to a MV Analysis

To calculate the mean variance frontier we need to calculate both the expected return and variance of different portfolios (based on n assets). With two assets ($n = 2$) the expected return and the variance of the portfolio are

$$\begin{aligned} \mathrm{E}\, R_p &= \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \\ \mathrm{Var}(R_p) &= \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}. \end{aligned} \tag{3.1}$$

In this case we need information on 2 mean returns and 3 elements of the covariance matrix. Clearly, the covariance matrix can alternatively be expressed as

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}, \tag{3.2}$$

which involves two variances and one correlation (as before, 3 elements).

There are two main problems in estimating these parameters: the number of parameters increase very quickly as the number of assets increases and historical estimates have proved to be somewhat unreliable for future periods.

To illustrate the first problem, notice that with n assets we need the following number of parameters

	Required number of estimates	With 100 assets
μ_i	n	100
σ_{ii}	n	100
σ_{ij}	$n(n-1)/2$	4950

The numerics is not the problem as it is a matter of seconds to estimate a covariance matrix of 100 return series. Instead, the problem is that most portfolio analysis uses lots of judgemental “estimates.” These are necessary since there might be new assets (no historical returns series are available) or there might be good reasons to believe that old estimates are not valid anymore. To cut down on the number of parameters, it is often assumed that returns follow some simple model. These notes will discuss so-called single- and multi-index models.

The second problem comes from the empirical observations that estimates from historical data are sometimes poor “forecasts” of future periods (which is what matters for portfolio choice). As an example, the correlation between two asset returns tends to be more “average” than the historical estimate would suggest.

A simple (and often used) way to deal with this is to replace the historical correlation with an average historical correlation. For instance, suppose there are three assets. Then, estimate ρ_{ij} on historical data, but use the average estimate as the “forecast” of all correlations:

$$\text{estimate} \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ & 1 & \rho_{23} \\ & & 1 \end{bmatrix}, \text{calculate } \bar{\rho} = (\hat{\rho}_{12} + \hat{\rho}_{13} + \hat{\rho}_{23})/3, \text{ and use } \begin{bmatrix} 1 & \bar{\rho} & \bar{\rho} \\ & 1 & \bar{\rho} \\ & & 1 \end{bmatrix}.$$

3.2 Single-Index Models

The single-index model is a way to cut down on the number of parameters that we need to estimate in order to construct the covariance matrix of assets. The model assumes that the co-movement between assets is due to a single common influence (here denoted R_m)

$$\begin{aligned} R_i &= \alpha_i + \beta_i R_m + e_i, \text{ where} \\ \mathrm{E}\, e_i &= 0,\ \mathrm{Cov}\,(e_i, R_m) = 0, \text{ and } \mathrm{Cov}(e_i, e_j) = 0. \end{aligned} \tag{3.3}$$

The first two assumptions are the standard assumptions for using Least Squares: the residual has a zero mean and is uncorrelated with the non-constant regressor. (Together they imply that the residuals are orthogonal to both regressors, which is the standard assumption in econometrics.) Hence, these two properties will be automatically satisfied if (3.3) is estimated by Least Squares.

See Figures 3.1 – 3.3 for illustrations.

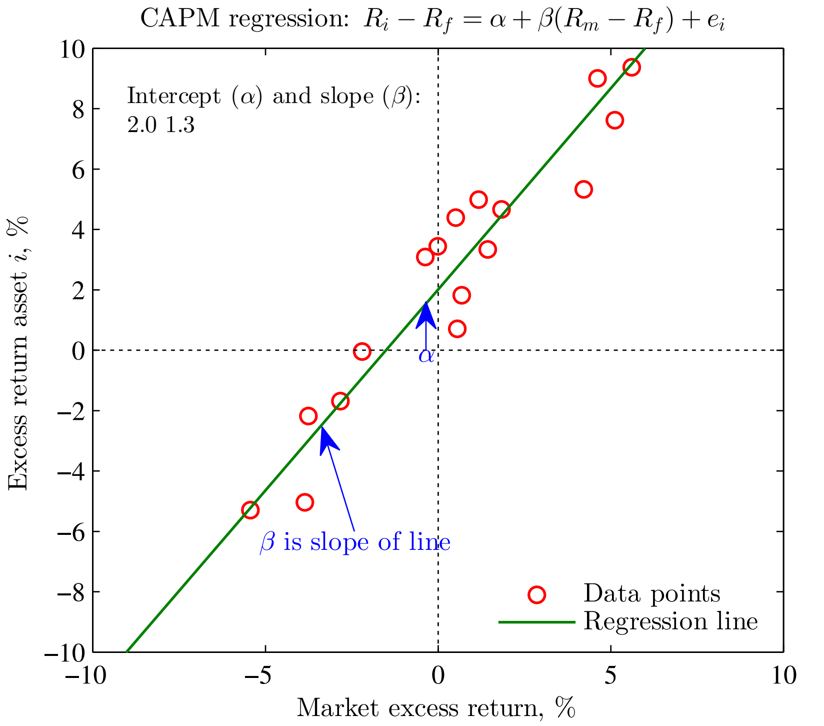


Figure 3.1: CAPM regression

The key point of the model, however, is the third assumption: the residuals for different assets are uncorrelated. This means that all comovements of two assets (R_i and R_j, say) are due to movements in the common "index" R_m. This is not at all guaranteed by running LS regressions—just an assumption. It is likely to be false—but may be a reasonable approximation in many cases. In any case, it simplifies the construction of the covariance matrix of the assets enormously—as demonstrated below.

Remark 3.1 *(The market model) The market model is (3.3) without the assumption that* $\text{Cov}(e_i, e_j) = 0$. *This model does not simplify the calculation of a portfolio variance—but will turn out to be important when we want to test CAPM.*

If (3.3) is true, then the variance of asset i and the covariance of assets i and j are

$$\sigma_{ii} = \beta_i^2 \text{ Var}(R_m) + \text{Var}(e_i) \tag{3.4}$$

$$\sigma_{ij} = \beta_i \beta_j \text{ Var}(R_m). \tag{3.5}$$

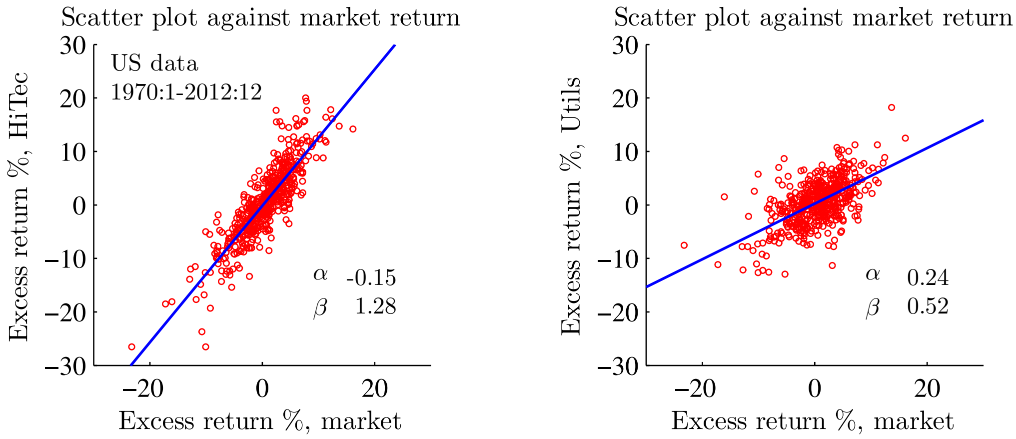


Figure 3.2: Scatter plot against market return

Together, these equations show that we can calculate the whole covariance matrix by having just the variance of the index (to get $\text{Var}(R_m)$) and the output from n regressions (to get β_i and $\text{Var}(e_i)$ for each asset). This is, in many cases, much easier to obtain than direct estimates of the covariance matrix. For instance, a new asset does not have a return history, but it may be possible to make intelligent guesses about its beta and residual variance (for instance, from knowing the industry and size of the firm).

This gives the covariance matrix (for two assets)

$$\text{Cov}\left(\begin{bmatrix} R_i \\ R_j \end{bmatrix}\right) = \begin{bmatrix} \beta_i^2 & \beta_i\beta_j \\ \beta_i\beta_j & \beta_j^2 \end{bmatrix} \text{Var}(R_m) + \begin{bmatrix} \text{Var}(e_i) & 0 \\ 0 & \text{Var}(e_j) \end{bmatrix}, \text{ or} \tag{3.6}$$

$$= \begin{bmatrix} \beta_i \\ \beta_j \end{bmatrix} \begin{bmatrix} \beta_i & \beta_j \end{bmatrix} \text{Var}(R_m) + \begin{bmatrix} \text{Var}(e_i) & 0 \\ 0 & \text{Var}(e_j) \end{bmatrix} \tag{3.7}$$

More generally, with n assets we can define β to be an $n \times 1$ vector of all the betas and Σ to be an $n \times n$ matrix with the variances of the residuals along the diagonal. We can then write the covariance matrix of the $n \times 1$ vector of the returns as

$$\text{Cov}(R) = \beta\beta' \, \text{Var}(R_m) + \Sigma. \tag{3.8}$$

See Figure 3.4 for an example based on the Fama-French portfolios detailed in Table 3.2.

	HiTec	Utils
constant	−0.15	0.24
	(−1.00)	(1.58)
market return	1.28	0.52
	(33.58)	(12.77)
R2	0.75	0.34
obs	516.00	516.00
Autocorr (t)	−0.73	0.86
White	6.19	20.42
All slopes	386.67	176.89

Table 3.1: CAPM regressions, monthly returns, %, US data 1970:1-2012:12. Numbers in parentheses are t-stats. Autocorr is a N(0,1) test statistic (autocorrelation); White is a chi-square test statistic (heteroskedasticity), df = K(K+1)/2 - 1; All slopes is a chi-square test statistic (of all slope coeffs), df = K-1

Remark 3.2 *(Fama-French portfolios) The portfolios in Table 3.2 are calculated by annual rebalancing (June/July). The US stock market is divided into 5×5 portfolios as follows. First, split up the stock market into 5 groups based on the book value/market value: put the lowest 20% in the first group, the next 20% in the second group etc. Second, split up the stock market into 5 groups based on size: put the smallest 20% in the first group etc. Then, form portfolios based on the intersections of these groups. For instance, in Table 3.2 the portfolio in row 2, column 3 (portfolio 8) belong to the 20%-40% largest firms and the 40%-60% firms with the highest book value/market value.*

	Book value/Market value				
	1	2	3	4	5
Size 1	1	2	3	4	5
2	6	7	8	9	10
3	11	12	13	14	15
4	16	17	18	19	20
5	21	22	23	24	25

Table 3.2: Numbering of the FF indices in the figures.

Proof. (of (3.4)–(3.5) By using (3.3) and recalling that $\text{Cov}(R_m, e_i) = 0$ direct calcu-

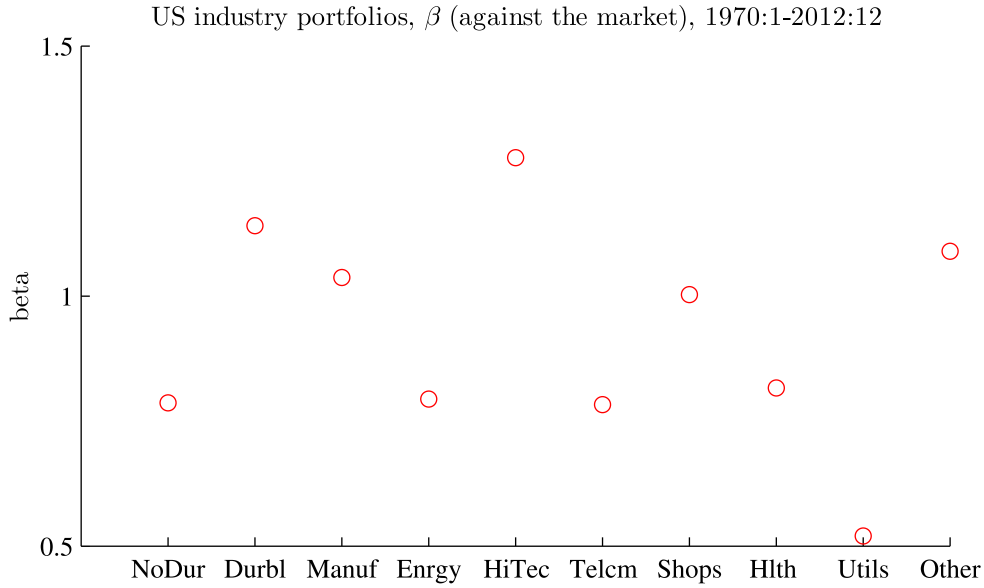


Figure 3.3: βs of US industry portfolios

lations give

$$\begin{aligned}
\sigma_{ii} &= \text{Var}\,(R_i) \\
&= \text{Var}\,(\alpha_i + \beta_i R_m + e_i) \\
&= \text{Var}\,(\beta_i R_m) + \text{Var}\,(e_i) + 2 \times 0 \\
&= \beta_i^2\, \text{Var}\,(R_m) + \text{Var}\,(e_i)\,.
\end{aligned}$$

Similarly, the covariance of assets i and j is (recalling also that $\text{Cov}\,\left(e_i, e_j\right) = 0$)

$$\begin{aligned}
\sigma_{ij} &= \text{Cov}\,\left(R_i, R_j\right) \\
&= \text{Cov}\,\left(\alpha_i + \beta_i R_m + e_i, \alpha_j + \beta_j R_m + e_j\right) \\
&= \beta_i \beta_j\, \text{Var}\,(R_m) + 0 \\
&= \beta_i \beta_j\, \text{Var}\,(R_m)\,.
\end{aligned}$$

■

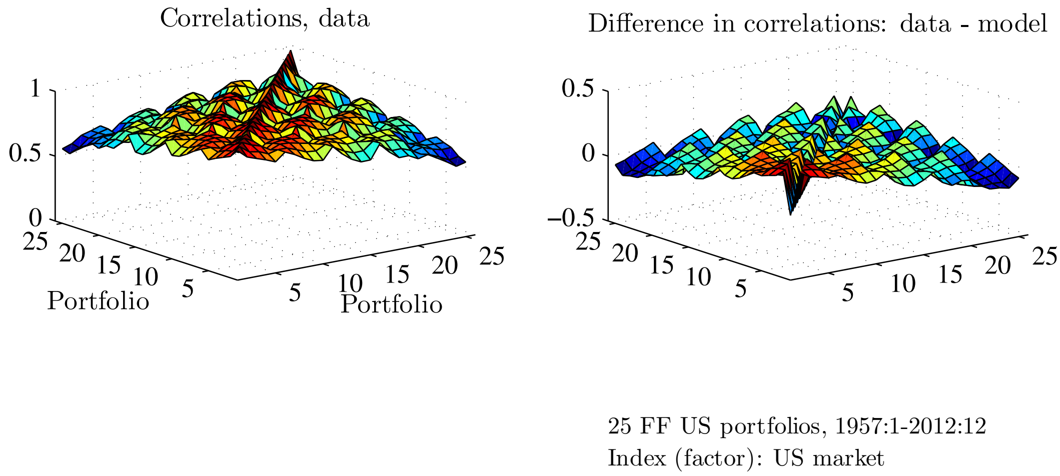


Figure 3.4: Correlations of US portfolios

3.3 Estimating Beta

3.3.1 Estimating Historical Beta: OLS and Other Approaches

Least Squares (LS) is typically used to estimate α_i, β_i and $\text{Std}(e_i)$ in (3.3)—and the R^2 is used to assess the quality of the regression.

Remark 3.3 *(R^2 of market model) R^2 of (3.3) measures the fraction of the variance (of R_i) that is due to the systematic part of the regression, that is, relative importance of market risk as compared to idiosyncratic noise ($1 - R^2$ is the fraction due to the idiosyncratic noise)*

$$R^2 = \frac{\text{Var}(\alpha_i + \beta_i R_m)}{\text{Var}(R_i)} = \frac{\beta_i^2 \sigma_m^2}{\beta_i^2 \sigma_m^2 + \sigma_{ei}^2}.$$

To assess the accuracy of historical betas, Blume (1971) and others estimate betas for non-overlapping samples (periods)—and then compare the betas across samples. They find that the correlation of betas across samples is moderate for individual assets, but relatively high for diversified portfolios. It is also found that betas tend to "regress" towards one: an extreme (high or low) historical beta is likely to be followed by a beta that is closer to one. There are several suggestions for how to deal with this problem.

To use *Blume's ad-hoc technique*, let $\hat{\beta}_{i1}$ be the estimate of β_i from an early sample, and $\hat{\beta}_{i2}$ the estimate from a later sample. Then regress

$$\hat{\beta}_{i2} = \gamma_0 + \gamma_1 \hat{\beta}_{i1} + \upsilon_i \tag{3.9}$$

and use it for forecasting the beta for yet another sample. Blume found $(\hat{\gamma}_0, \hat{\gamma}_1) = (0.343, 0.677)$ in his sample.

Other authors have suggested averaging the OLS estimate $(\hat{\beta}_{i1})$ with some average beta. For instance, $(\hat{\beta}_{i1}+1)/2$ (since the average beta must be unity) or $(\hat{\beta}_{i1}+\Sigma_{i=1}^{n}\hat{\beta}_{i1}/n)/2$ (which will typically be similar since $\Sigma_{i=1}^{n}\hat{\beta}_{i1}/n$ is likely to be close to one).

The *Bayesian approach* is another (more formal) way of adjusting the OLS estimate. It also uses a weighted average of the OLS estimate, $\hat{\beta}_{i1}$, and some other number, β_0, $(1-F)\hat{\beta}_{i1} + F\beta_0$ where F depends on the precision of the OLS estimator. The general idea of a Bayesian approach (Greene (2003) 16) is to treat both R_i and β_i as random. In this case a Bayesian analysis could go as follows. First, suppose our prior beliefs (before having data) about β_i is that it is normally distributed, $N(\beta_0, \sigma_0^2)$, where (β_0, σ_0^2) are some numbers . Second, run a LS regression of (3.3). If the residuals are normally distributed, so is the estimator—it is $N(\hat{\beta}_{i1}, \sigma_{\beta 1}^2)$, where we have taken the point estimate to be the mean. If we treat the variance of the LS estimator ($\sigma_{\beta 1}^2$) as known, then the Bayesian estimator of beta is

$$\begin{aligned} b &= (1-F)\hat{\beta}_{i1} + F\beta_0, \text{ where} \\ F &= \frac{1/\sigma_0^2}{1/\sigma_0^2 + 1/\sigma_{\beta 1}^2} = \frac{\sigma_{\beta 1}^2}{\sigma_0^2 + \sigma_{\beta 1}^2}. \end{aligned} \tag{3.10}$$

When the prior beliefs are very precise ($\sigma_0^2 \to 0$), then $F \to 1$ so the Bayesian estimator is the same as the prior mean. Effectively, when the prior beliefs are so precise, there is no room for data to add any information. In contrast, when the prior beliefs are very imprecise ($\sigma_0^2 \to \infty$), then $F \to 0$, so the Bayesian estimator is the same as OLS. Effectively, the prior beliefs do not add any information. In the current setting, $\beta_0 = 1$ and σ_0^2 taken from a previous (econometric) study might make sense.

3.3.2 Fundamental Betas

Another way to improve the forecasts of the beta over a future period is to bring in information about fundamental firm variables. This is particularly useful when there is little historical data on returns (for instance, because the asset was not traded before).

It is often found that betas are related to fundamental variables as follows (with signs in parentheses indicating the effect on the beta): Dividend payout (-), Asset growth (+), Leverage (+), Liquidity (-), Asset size (-), Earning variability (+), Earnings Beta (slope in earnings regressed on economy wide earnings) (+). Such relations can be used to make an educated guess about the beta of an asset without historical data on the returns—but with data on (at least some) of these fundamental variables.

3.4 Multi-Index Models

3.4.1 Overview

The multi-index model is just a multivariate extension of the single-index model (3.3)

$$\begin{aligned} R_i &= a_i^* + \textstyle\sum_{k=1}^{K} b_{ik}^* I_k^* + e_i, \text{ where} \\ \mathrm{E}\, e_i &= 0,\ \mathrm{Cov}\left(e_i, I_k^*\right) = 0, \text{ and } \mathrm{Cov}(e_i, e_j) = 0. \end{aligned} \tag{3.11}$$

As an example, there could be two indices: the stock market return and an interest rate. An ad-hoc approach is to first try a single-index model and then test if the residuals are approximately uncorrelated. If not, then adding a second index might improve the model.

It is often found that it takes several indices to get a reasonable approximation—but that a single-index model is equally good (or better) at "forecasting" the covariance over a future period. This is much like the classical trade-off between in-sample fit (requires a large model) and forecasting (often better with a small model).

The types of indices vary, but one common set captures the "business cycle" and includes things like the market return, interest rate (or some measure of the yield curve slope), GDP growth, inflation, and so forth. Another common set of indices are industry indices.

It turns out (see below) that the calculations of the covariance matrix are much simpler if the indices are transformed to be uncorrelated so we get the model

$$
\begin{aligned}
R_i &= a_i + \textstyle\sum_{k=1}^{K} b_{ik} I_k + e_i, \text{ where} \qquad (3.12)\\
\mathrm{E}\, e_i &= 0,\ \mathrm{Cov}\,(e_i, I_k) = 0,\ \mathrm{Cov}(e_i, e_j) = 0 \text{ (unless } i = j\text{), and}\\
\mathrm{Cov}(I_k, I_h) &= 0 \text{ (unless } k = h\text{)}.
\end{aligned}
$$

If this transformation of the indices is linear (and non-singular, so it is can be reversed if we want to), then the fit of the regression is unchanged.

3.4.2 "Rotating" the Indices

There are several ways of transforming the indices to make them uncorrelated, but the following regression approach is perhaps the simplest and may also give the best possibility of interpreting the results:

1. Let the first transformed index equal the original index, $I_1 = I_1^*$ (possibly demeaned). This would often be the market return.

2. Regress the second original index on the first transformed index, $I_2^* = \gamma_0 + \gamma_1 I_1 + \varepsilon_2$. Then, let the second transformed index be the fitted residual, $I_2 = \gamma_0 + \hat{\varepsilon}_2$.

3. Regress the third original index on the first two transformed indices, $I_3^* = \theta_0 + \theta_1 I_1 + \theta_2 I_2 + \varepsilon_3$. Then, let $I_3 = \theta_0 + \hat{\varepsilon}_3$. Follow the same idea for all subsequent indices.

Recall that the fitted residual (from Least Squares) is always uncorrelated with the regressor (by construction). In this case, this means that I_2 is uncorrelated with I_1 (step 2) and that I_3 is uncorrelated with both I_1 and I_2 (step 3). The correlation matrix of the first three rotated indices is therefore

$$
\mathrm{Corr}\left(\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \qquad (3.13)
$$

This recursive approach also helps in interpreting the transformed indices. Suppose the first index is the market return and that the second original index is an interest rate. The first transformed index (I_1) is then clearly the market return. The second transformed index (I_2) can then be interpreted as the interest rate minus the interest rate expected at the current stock market return—that is, the part of the interest rate that cannot be explained by the stock market return.

More generally, let the kth index ($k = 1, 2, \ldots, K$) be

$$I_k = \delta_{k1} + \hat{\varepsilon}_k, \tag{3.14}$$

where δ_{k1} and $\hat{\varepsilon}_k$ are the fitted interecept and residual from the regression

$$I_k^* = \delta_{k1} + \textstyle\sum_{s=1}^{k-1} \gamma_{ks} I_s + \varepsilon_k. \tag{3.15}$$

Notice that for the first index ($k = 1$), the regression is only $I_1^* = \delta_{11} + \varepsilon_1$, so I_1 equals I_1^*.

3.4.3 Multi-Index Model after "Rotating" the Indices

To see why the transformed indices are very convenient for calculating the covariance matrix, consider a two-index model. Then, (3.12) implies that the variance of asset i is

$$\begin{aligned}\sigma_{ii} &= \operatorname{Var}(a_i + b_{i1} I_1 + b_{i2} I_2 + e_i)\\ &= b_{i1}^2 \operatorname{Var}(I_1) + b_{i2}^2 \operatorname{Var}(I_2) + \operatorname{Var}(e_i).\end{aligned} \tag{3.16}$$

Similarly, the covariance of assets i and j is

$$\begin{aligned}\sigma_{ij} &= \operatorname{Cov}\left(a_i + b_{i1} I_1 + b_{i2} I_2 + e_i, a_j + b_{j1} I_1 + b_{j2} I_2 + e_j\right)\\ &= b_{i1} b_{j1} \operatorname{Var}(I_1) + b_{i2} b_{j2} \operatorname{Var}(I_2).\end{aligned} \tag{3.17}$$

More generally, with n assets and K indices we can define b_1 to be an $n \times 1$ vector of the slope coefficients for the first index (b_{i1}, b_{j1}) and b_2 the vector of slope coefficients for the second index and so on. Also, let Σ to be an $n \times n$ matrix with the variances of the residuals along the diagonal. The covariance matrix of the returns is then

$$\operatorname{Cov}(R) = b_1 b_1' \operatorname{Var}(I_1) + b_2 b_2' \operatorname{Var}(I_2) + \ldots + b_K b_K' \operatorname{Var}(I_K) + \Sigma \tag{3.18}$$

$$= \textstyle\sum_{k=1}^{K} b_k b_k' \operatorname{Var}(I_k) + \Sigma. \tag{3.19}$$

See Figure 3.5 for an example.

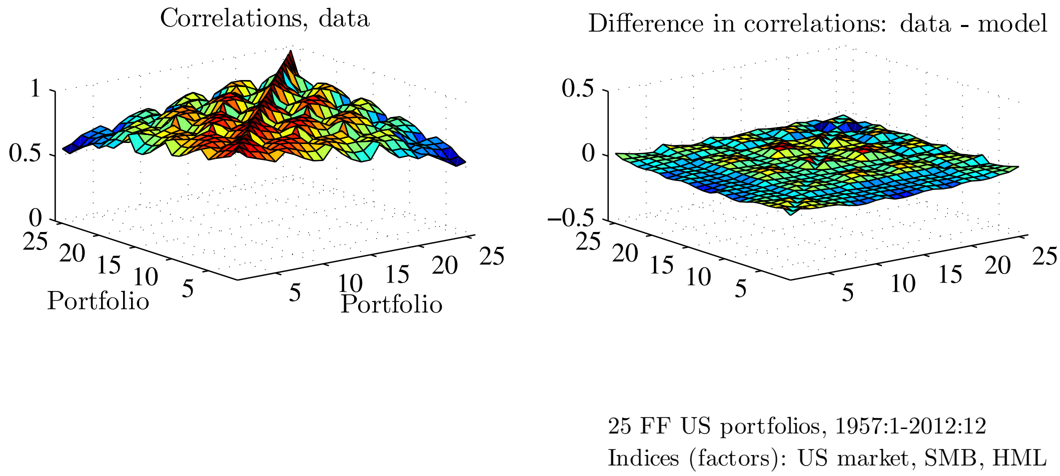


Figure 3.5: Correlations of US portfolios

3.4.4 Multi-Index Model as a Method for Portfolio Choice

The factor loadings (betas) can be used for more than just constructing the covariance matrix. In fact, the factor loadings are often used directly in portfolio choice. The reason is simple: the betas summarize how different assets are exposed to the big risk factors/return drivers. The betas therefore provide a way to understand the broad features of even complicated portfolios. Combined this with the fact that many analysts and investors have fairly little direct information about individual assets, but are often willing to form opinions about the future relative performance of different asset classes (small vs large firms, equity vs bonds, etc)—and the role for factor loadings becomes clear.

See Figures 3.6–3.7 for an illustration.

3.5 Estimating Expected Returns

The starting point for forming estimates of future mean excess returns is typically historical excess returns. Excess returns are preferred to returns, since this avoids blurring the risk compensation (expected excess return) with long-run movements in inflation (and therefore interest rates). The expected excess return for the future period is typically formed as a judgmental adjustment of the historical excess return. Evidence suggest that

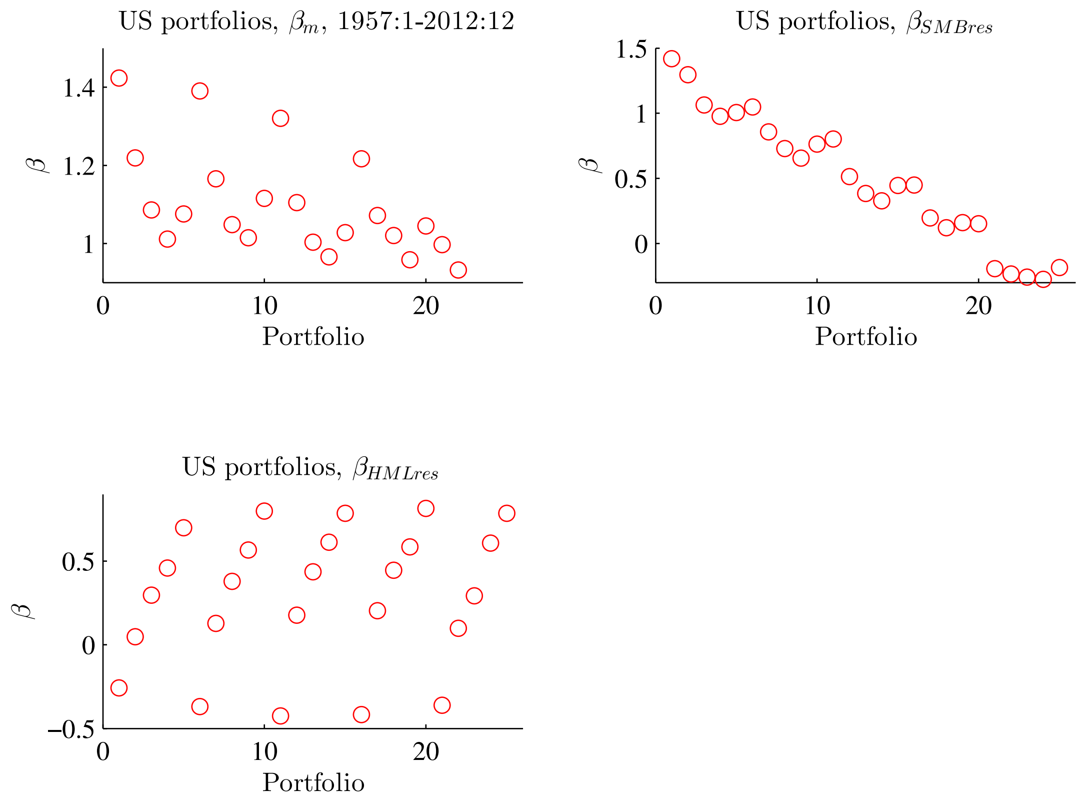


Figure 3.6: Loading (betas) of rotated factors

the adjustments are hard to make.

It is typically hard to predict movements (around the mean) of asset returns, but a few variables seem to have some predictive power, for instance, the slope of the yield curve, the earnings/price yield, and the book value–market value ratio. Still, the predictive power is typically low.

Makridakis, Wheelwright, and Hyndman (1998) 10.1 show that there is little evidence that the average stock analyst beats (on average) the market (a passive index portfolio). In fact, less than half of the analysts beat the market. However, there are analysts which seem to outperform the market for some time, but the autocorrelation in over-performance is weak. The evidence from mutual funds is similar. For them it is typically also found that their portfolio weights do not anticipate price movements.

It should be remembered that many analysts also are sales persons: either of a stock (for instance, since the bank is underwriting an offering) or of trading services. It could

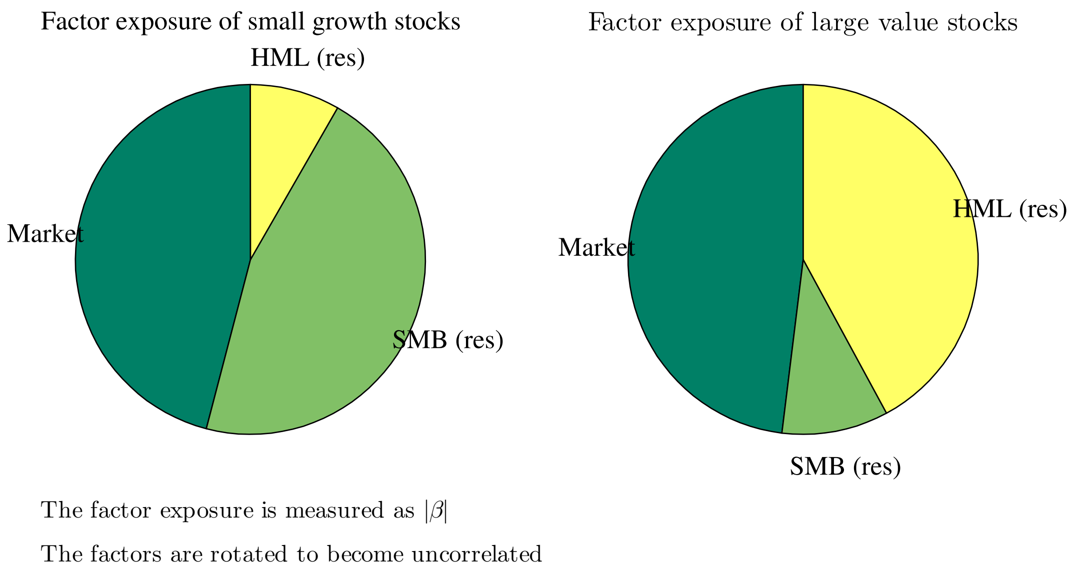


Figure 3.7: Absolute loading (betas) of rotated factors

well be that their objective function is quite different from minimizing the squared forecast errors—or whatever we typically use in order to evaluate their performance. (The number of litigations in the US after the technology boom/bust should serve as a strong reminder of this.)

Bibliography

Amemiya, T., 1985, *Advanced econometrics*, Harvard University Press, Cambridge, Massachusetts.

Blume, M. E., 1971, "On the Assessment of Risk," *Journal of Finance*, 26, 1–10.

Elton, E. J., M. J. Gruber, S. J. Brown, and W. N. Goetzmann, 2010, *Modern portfolio theory and investment analysis*, John Wiley and Sons, 8th edn.

Greene, W. H., 2003, *Econometric analysis*, Prentice-Hall, Upper Saddle River, New Jersey, 5th edn.

Makridakis, S., S. C. Wheelwright, and R. J. Hyndman, 1998, *Forecasting: methods and applications*, Wiley, New York, 3rd edn.

4 Risk Measures

Reference: Hull (2006) 18; McDonald (2006) 25; Fabozzi, Focardi, and Kolm (2006) 4–5; McNeil, Frey, and Embrechts (2005); Alexander (2008)

4.1 Symmetric Dispersion Measures

4.1.1 Mean Absolute Deviation

The variance (and standard deviation) is very sensitive to the tails of the distribution. For instance, even if the standard normal distribution and a student-t distribution with 4 degrees of freedom look fairly similar, the latter has a variance that is twice as large (recall: the variance of a t_n distribution is $n/(n-2)$ for $n > 2$). This may or may not be what the investor cares about. If not, the mean absolute deviation is an alternative. Let μ be the mean, then the definition is

$$\text{mean absolute deviation} = \mathrm{E}\,|R - \mu|. \tag{4.1}$$

This measure of dispersion is much less sensitive to the tails—essentially because it does not involve squaring the variable.

Notice, however, that for a normally distributed return the mean absolute deviation is proportional to the standard deviation—see Remark 4.1. Both measures will therefore lead to the same portfolio choice (for a given mean return). In other cases, the portfolio choice will be different (and perhaps complicated to perform since it is typically not easy to calculate the mean absolute deviation of a portfolio).

Remark 4.1 *(Mean absolute deviation of $N(\mu, \sigma^2)$ and t_n) If $R \sim N(\mu, \sigma^2)$, then*

$$\mathrm{E}\,|R - \mu| = \sqrt{2/\pi}\sigma \approx 0.8\sigma.$$

If $R \sim t_n$, then $\mathrm{E}\,|R| = 2\sqrt{n}/[(n-1)B(n/2, 0.5)]$, where B is the beta function. For $n = 4$, $\mathrm{E}\,|R| = 1$ which is just 25% higher than for a $N(0,1)$ distribution. In contrast, the standard deviation is $\sqrt{2}$, which is 41% higher than for the $N(0,1)$.

4.1.2 Index Tracking Errors

Suppose instead that our task, as fund managers, say, is to track a benchmark portfolio (returns R_b and portfolio weights w_b)—but we are allowed to make some deviations. For instance, we are perhaps asked to track a certain index. The deviations, typically measured in terms of the variance of the tracking errors for the returns, can be motivated by practical considerations and by concerns about trading costs. If our portfolio has the weights w, then the portfolio return is $R_p = w'R$, where R are the original assets. Similarly, the benchmark portfolio (index) has the return $R_b = w_b'R$. If the variance of the tracking error should be less than U, then we have the restriction

$$\text{Var}(R_p - R_b) = (w - w_b)'\Sigma(w - w_b) \leq U, \tag{4.2}$$

where Σ is the covariance matrix of the original assets. This type of restriction is fairly easy to implement numerically in the portfolio choice model (the optimization problem).

4.2 Downside Risk

4.2.1 Value at Risk

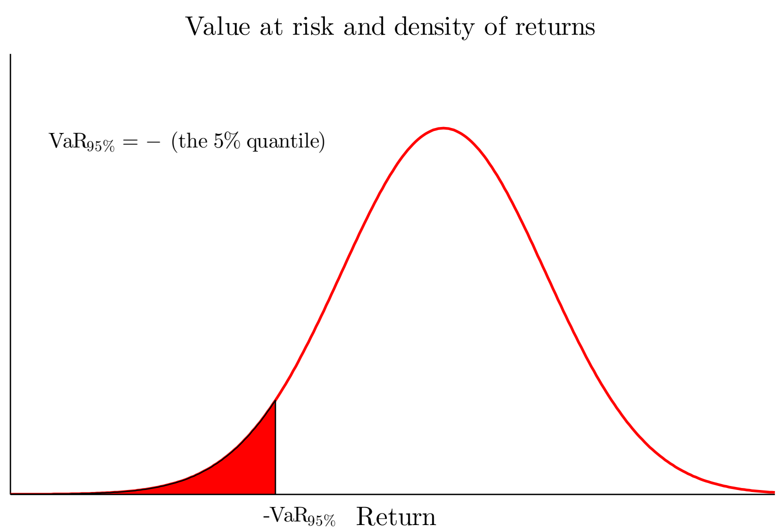


Figure 4.1: Value at risk

The mean-variance framework is often criticized for failing to distinguish between downside (considered to be risk) and upside (considered to be potential).

The 95% Value at Risk ($\text{VaR}_{95\%}$) is a number such that there is only a 5% chance that the loss ($-R$) is larger that $\text{VaR}_{95\%}$

$$\Pr(\text{Loss} \geq \text{VaR}_{95\%}) = \Pr(-R \geq \text{VaR}_{95\%}) = 5\%. \tag{4.3}$$

Here, 95% is the confidence level of the VaR. Clearly, $-R \geq \text{VaR}_{95\%}$ is true when (and only when) $R \leq -\text{VaR}_{95\%}$, so (4.3) can also be expressed as

$$\Pr(R \leq -\text{VaR}_{95\%}) = \text{cdf}_R(-\text{VaR}_{95\%}) = 5\%, \tag{4.4}$$

where $\text{cdf}_R()$ is the cumulative distribution function of the returns. This says that $-\text{VaR}_{95\%}$ is a number such that there is only a 5% chance that the return is below it. That is, $-\text{VaR}_\alpha$ is the 0.05 quantile (5th percentile) of the return distribution. Using (4.4) allows us to work directly with the return distribution (not the loss distribution), which is often convenient. See Figures 4.1–4.2 for illustrations.

Example 4.2 *(Quantile of a distribution) The 0.05 quantile is the value such that there is only a 5% probability of a lower number,* $\Pr(R \leq \textit{quantile}_{0.05}) = 0.05$.

This can be expressed more formally by solving (4.4) for the value at risk, $\text{VaR}_{95\%}$, as

$$\text{VaR}_{95\%} = -\,\text{cdf}_R^{-1}(0.05), \tag{4.5}$$

where $\text{cdf}_R^{-1}()$ is the inverse of the cumulative distribution function of the returns, so $\text{cdf}_R^{-1}(0.05)$ is the 0.05 quantile (or "critical value") of the return distribution. To convert the value at risk into value terms (CHF, say), just multiply the VaR for returns with the value of the investment (portfolio). If the return is normally distributed, $R \sim N(\mu, \sigma^2)$ then

$$\text{VaR}_{95\%} = -(\mu - 1.64\sigma). \tag{4.6}$$

More generally, a there is only a $1-\alpha$ chance that the loss ($-R$) is larger that VaR_α (the confidence level is α)

$$\Pr(-R \geq \text{VaR}_\alpha) = 1 - \alpha, \text{ so} \tag{4.7}$$

$$\text{VaR}_\alpha = -\,\text{cdf}_R^{-1}(1-\alpha). \tag{4.8}$$

If the return is normally distributed, $R \sim N(\mu, \sigma^2)$ and $c_{1-\alpha}$ is the $1-\alpha$ quantile of a N(0,1) distribution (for instance, -1.64 for $1-\alpha = 0.05$), then

$$\text{VaR}_{\alpha} = -(\mu + c_{1-\alpha}\sigma). \tag{4.9}$$

This is illustrated in Figure 4.4.

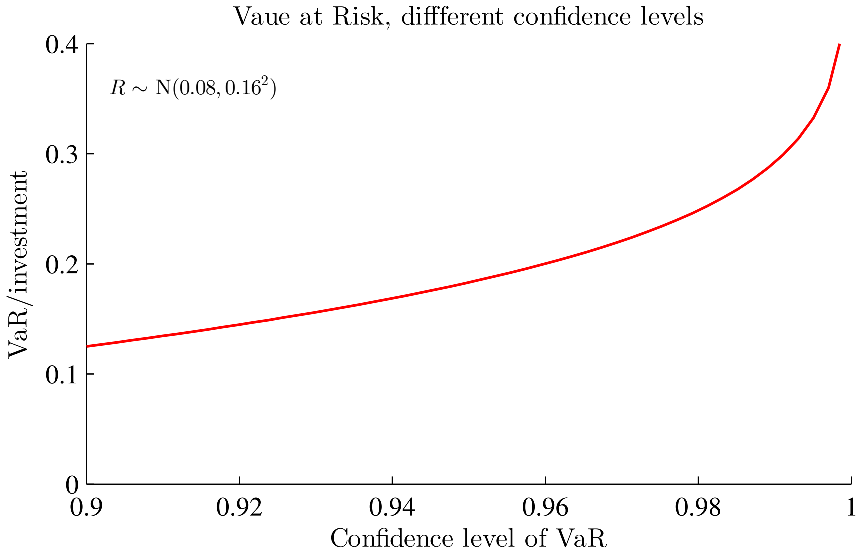


Figure 4.2: Value at risk, different probability levels

Remark 4.3 *(Critical values of $N(\mu, \sigma^2)$) If $R \sim N(\mu, \sigma^2)$, then there is a 5% probability that $R \leq \mu - 1.64\sigma$, a 2.5% probability that $R \leq \mu - 1.96\sigma$, and a 1% probability that $R \leq \mu - 2.33\sigma$.*

Example 4.4 *(VaR with $R \sim N(\mu, \sigma^2)$) If daily returns have $\mu = 8\%$ and $\sigma = 16\%$, then the 1-day $VaR_{95\%} = -(0.08 - 1.64 \times 0.16) \approx 0.18$; we are 95% sure that we will not loose more than 18% of the investment over one day, that is, $VaR_{95\%} = 0.18$. Similarly, $VaR_{97.5\%} = -(0.08 - 1.96 \times 0.16) \approx 0.24$.*

Figure 4.3 shows the distribution and VaRs (for different probability levels) for the daily S&P 500 returns. Two different VaRs are shown: based on a normal distribution and as the empirical VaR (from the empirical quantiles of the distribution). While these

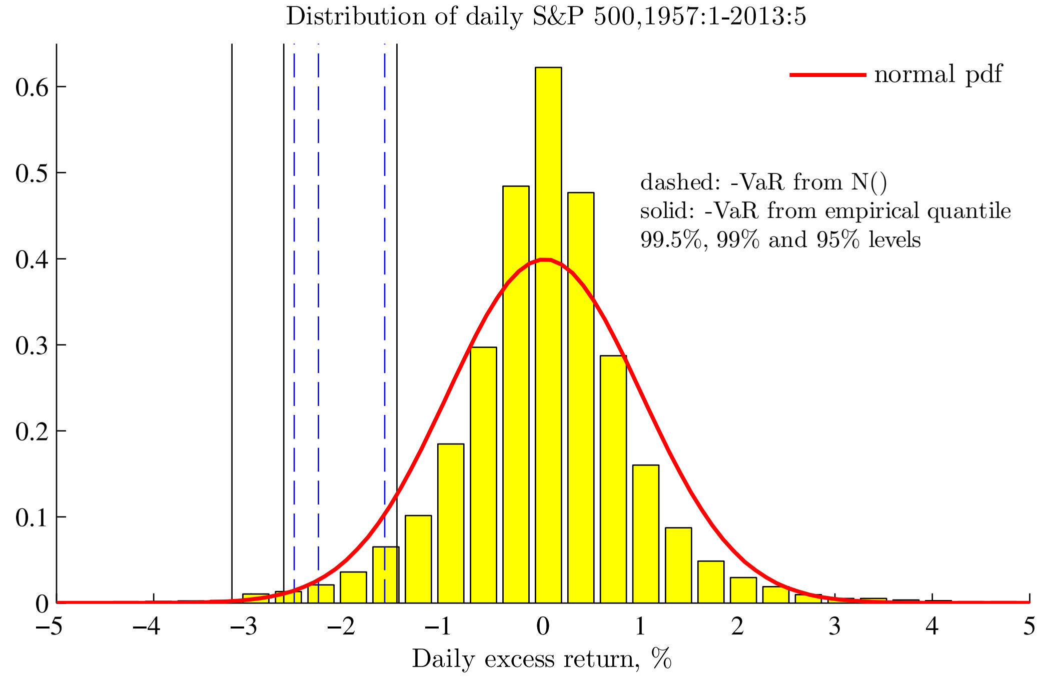


Figure 4.3: Return distribution and VaR for S&P 500

results are interesting, they are just time-averages in the sense of being calculated from the unconditional distribution: time-variation in the distribution is not accounted for.

Figure 4.5 illustrates the VaR calculated from a time series model (to be precise, an AR(1)+GARCH(1,1) model) for daily S&P returns. In this case, the VaR changes from day to day as both the mean return (the forecast) as well as the standard error (of the forecast error) do. Since *volatility clearly changes over time*, this is crucial for a reliable VaR model.

Notice that the value at risk in (4.9), that is, when the return is normally distributed, is a strictly increasing function of the standard deviation (and the variance). This follows from the fact that $c_{1-\alpha} < 0$ (provided $1-\alpha < 50\%$, which is the relevant case). Minimizing the VaR at a given mean return therefore gives the same solution (portfolio weights) as minimizing the variance at the same given mean return. In other cases, the portfolio choice will be different (and perhaps complicated to perform).

Example 4.5 *(VaR and regulation of bank capital) Bank regulations have used 3 times the 99% VaR for 10-day returns as the required bank capital.*

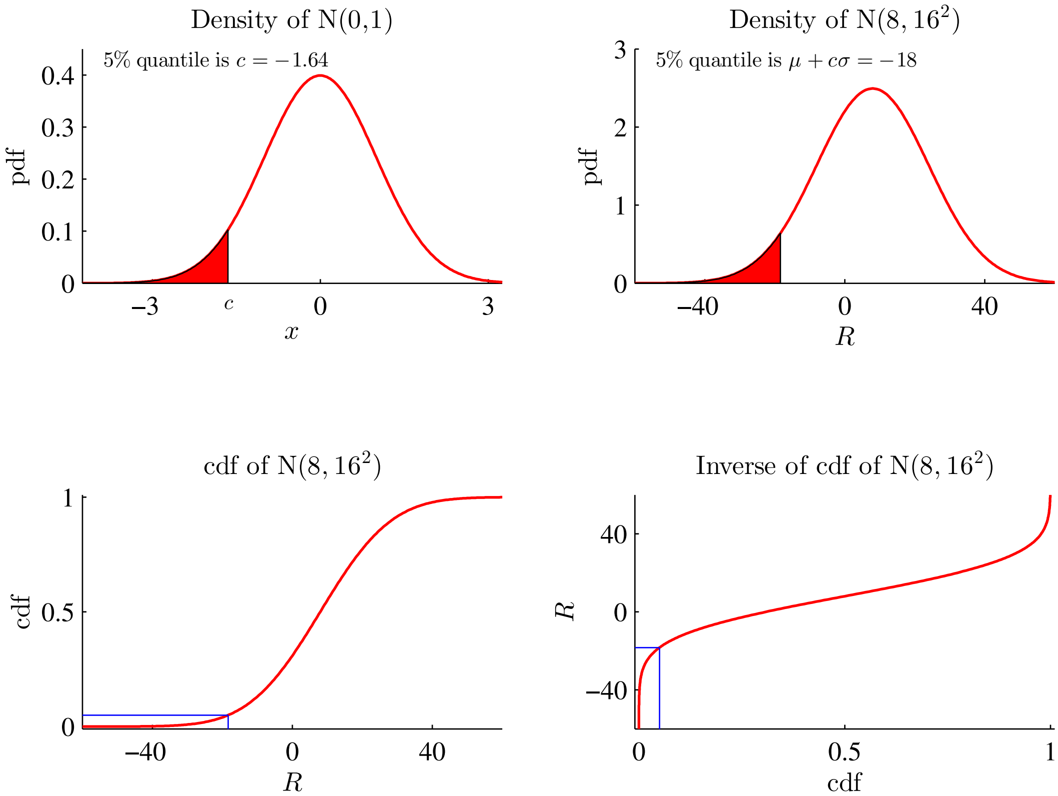


Figure 4.4: Finding critical value of N(μ,σ^2) distribution

Notice that the return distribution depends on the investment horizon, so a value at risk measure is typically calculated for a stated investment period (for instance, one day). Multi-period VaRs are calculated by either explicitly constructing the distribution of multi-period returns, or by making simplifying assumptions about the relation between returns in different periods (for instance, that they are iid).

Remark 4.6 *(Multi-period VaR) If the returns are iid, then a q-period return has the mean $q\mu$ and variance $q\sigma^2$, where μ and σ^2 are the mean and variance of the one-period returns respectively. If the mean is zero, then the q-day VaR is $\sqrt{q}$ times the one-day VaR.*

4.2.2 Backtesting a VaR model

Backtesting a VaR model amounts to checking if (historical) data fits with the VaR numbers. For instance, we first find the $\text{VaR}_{95\%}$ and then calculate what fraction of returns

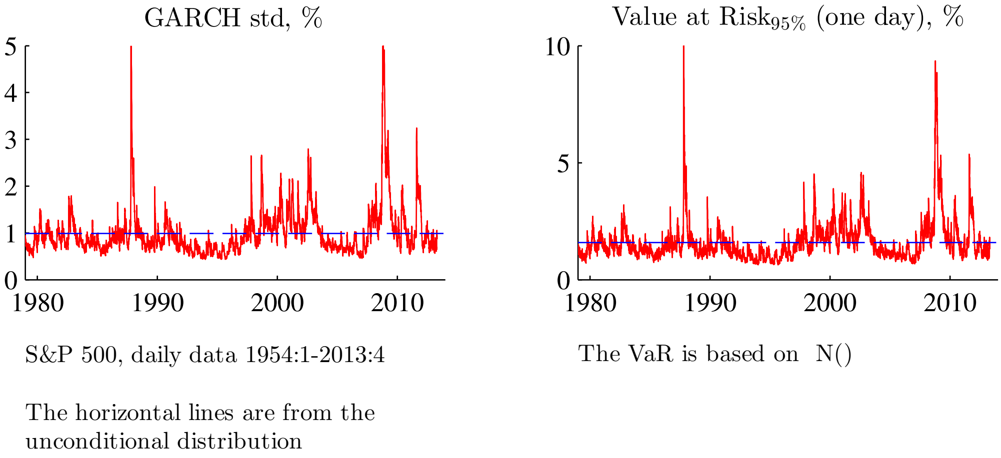


Figure 4.5: Conditional volatility and VaR

that is actually below (the negative of) this number. If the model is correct it should be 5%. We then repeat this for VaR$_{96\%}$—only 4% of the returns should be below (the negative of) this number. Figures 4.6–4.7 show results from backtesting a VaR model where the volatility follows a GARCH process. It suggests that a GARCH model (to capture the time varying volatility), combined with the assumption that the return is normally distributed (but with time-varying parameters), works relatively well.

The VaR concept has been criticized for having poor aggregation properties. In particular, the VaR for a portfolio is not necessarily (weakly) lower than the portfolio of the VaRs, which contradicts the notion of diversification benefits. (To get this unfortunate property, the return distributions must be heavily skewed.)

See Table 4.1 for an empirical comparison of the VaR with some alternative downside risk measures (discussed below).

	Small growth	Large value
Std	8.0	5.0
VaR (95%)	12.3	8.3
ES (95%)	17.2	10.8
SemiStd	5.5	3.4
Drawdown	79.7	52.3

Table 4.1: Risk measures of monthly returns of two stock indices (%), US data 1957:1-2012:12.

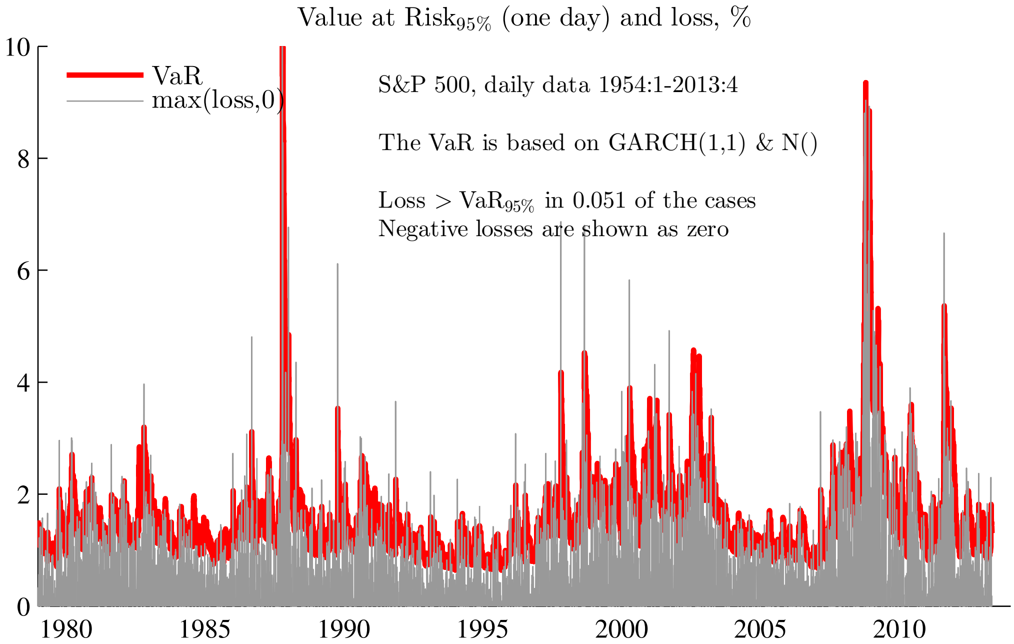


Figure 4.6: Backtesting VaR from a GARCH model, assuming normally distributed shocks

4.2.3 Value at Risk of a Portfolio

If the return distribution is normal with a zero mean, $R_i \sim N(0, \sigma_i^2)$, then the 95% value at risk for asset i is

$$\text{VaR}_i = 1.64\sigma_i. \tag{4.10}$$

(Warning: VaR_i now stands for the value at risk of asset i.) It is then straightfoward to show that the VaR for a portfortfolio

$$R_p = w_1 R_1 + w_2 R_2, \tag{4.11}$$

where $w_1 + w_2 = 1$ can be written

$$\text{VaR}_p = \left(\begin{bmatrix} w_1\text{VaR}_1 & w_2\text{VaR}_2 \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix} \begin{bmatrix} w_1\text{VaR}_1 \\ w_2\text{VaR}_2 \end{bmatrix} \right)^{1/2}, \tag{4.12}$$

where ρ_{12} is the correlation of R_1 and R_2. The extension to n (instead of 2) assets is straightforward.

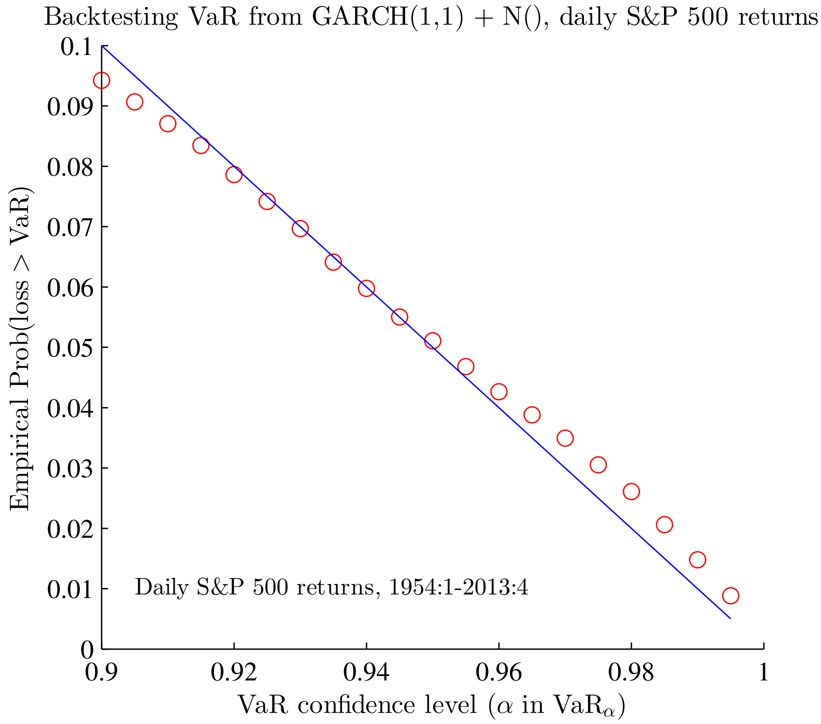


Figure 4.7: Backtesting VaR from a GARCH model, assuming normally distributed shocks

This expression highlights the importance of both the individual VaR_i values and the correlation. Clearly, a worst case scenario is when the portfolio is long in all assets ($w_i > 0$) and the correlation turns out to be perfect ($\rho_{12} = 1$). In this case, there is no diversification benefits so the portfolio variance is high—which leads to a high value at risk.

Proof. (of (4.12)) Recall that $\text{VaR}_p = 1.64\sigma_p$, and that

$$\sigma_p^2 = w_1^2\sigma_{11} + w_2^2\sigma_{22} + 2w_1w_2\rho_{12}\sigma_1\sigma_2.$$

Use (4.10) to substitute as $\sigma_i = \text{VaR}_i/1.64$

$$\sigma_p^2 = w_1^2\text{VaR}_1^2/1.64^2 + w_2^2\text{VaR}_2^2/1.64^2 + 2w_1w_2\rho_{12} \times \text{VaR}_1 \times \text{VaR}_2/1.64^2.$$

Multiply both sides by 1.64^2 and take the square root to get (4.12). ■

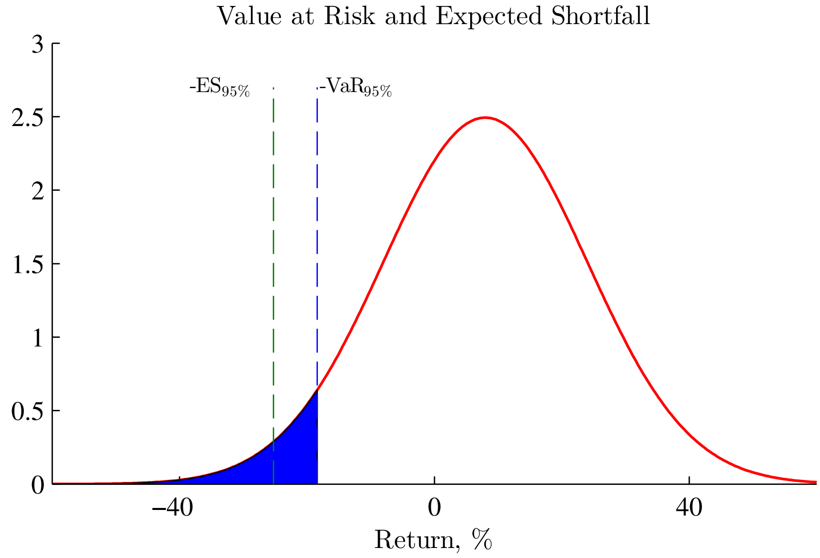


Figure 4.8: Value at risk and expected shortfall

4.2.4 Index Models for Calculating the Value at Risk

Consider a multi-index model

$$
\begin{aligned}
R &= a + b_1 I_1 + b_2 I_2 + \ldots + b_k I_k + e, \text{ or} \\
&= a + b' I + e,
\end{aligned}
\tag{4.13}
$$

where b is a $k \times 1$ vector of the b_i coefficients and I is also a $k \times 1$ vector of the I_i indices. As usual, we assume $\mathrm{E}\, e = 0$ and $\mathrm{Cov}\,(e, I_i) = 0$. This model can be used to generate the inputs to a VaR model. For instance, the mean and standard deviation of the return are

$$
\begin{aligned}
\mu &= a + b' \,\mathrm{E}\, I \\
\sigma &= \sqrt{b' \,\mathrm{Cov}(I) b + \mathrm{Var}(e)},
\end{aligned}
\tag{4.14}
$$

which can be used in (4.9), that is, an assumption of a normal return distribution. If the return is of a well diversified portfolio and the indices include the key stock indices, then the idiosyncratic risk $\mathrm{Var}(e)$ is close to zero. The RiskMetrics approach is to make this assumption.

Stand-alone VaR is a way to assess the contribution of different factors (indices). For instance, the indices in (4.13) could include: an equity indices, interest rates, exchange rates and perhaps also a few commodity indices. Then, an *equity VaR* is calculated by setting all elements in b, except those for the equity indices, to zero. Often, the intercept, a, is also set to zero. Similarly, an *interest rate VaR* is calculated by setting all elements in b, except referring to the interest rates, to zero. And so forth for an *FX VaR* and a *commodity VaR*. Clearly, these different VaRs do not add up to the total VaR, but they still give an indication of where the main risk comes from.

If an asset or a portfolio is a non-linear function of the indices, then (4.13) can be thought of as a first-order Taylor approximation where b_i represents the partial derivative of the asset return with respect to index i. For instance, an option is a non-linear function of the underlying asset value and its volatility (as well as the time to expiration and the interest rate). This approach, when combined with the normal assumption in (4.9), is called the *delta-normal method*.

4.2.5 VaR and Portfolio Choice

Consider the case of one risky asset (R_1) and a riskfree asset (R_f). If the portfolio weight on the risky asset is v, then the key properties of the portfolio are

$$\begin{aligned}
R_p &= vR_1 + (1-v)R_f \text{, so} \\
\mathrm{E}\, R_p &= v\, \mathrm{E}\, R_1 + (1-v)R_f \text{ and} \\
\mathrm{Std}(R_p) &= |v|\, \mathrm{Std}(R_1) \\
\mathrm{VaR}_{95\%} &= -[\mathrm{E}\, R_p - 1.64\, \mathrm{Std}(R_p)]. \qquad (4.15)
\end{aligned}$$

The effect of changing the portfolio weight is illustrated in Figure 4.9.

4.2.6 Expected Shortfall

The expected shortfall (also called conditional VaR, average value at risk and expected tail loss) is the expected loss when the return actually is below the VaR_α, that is,

$$\mathrm{ES}_\alpha = -\,\mathrm{E}(R | R \le -\mathrm{VaR}_\alpha). \qquad (4.16)$$

This might be more informative than the VaR_α, which is the *minimum loss* that will happen with a $1-\alpha$ probability.

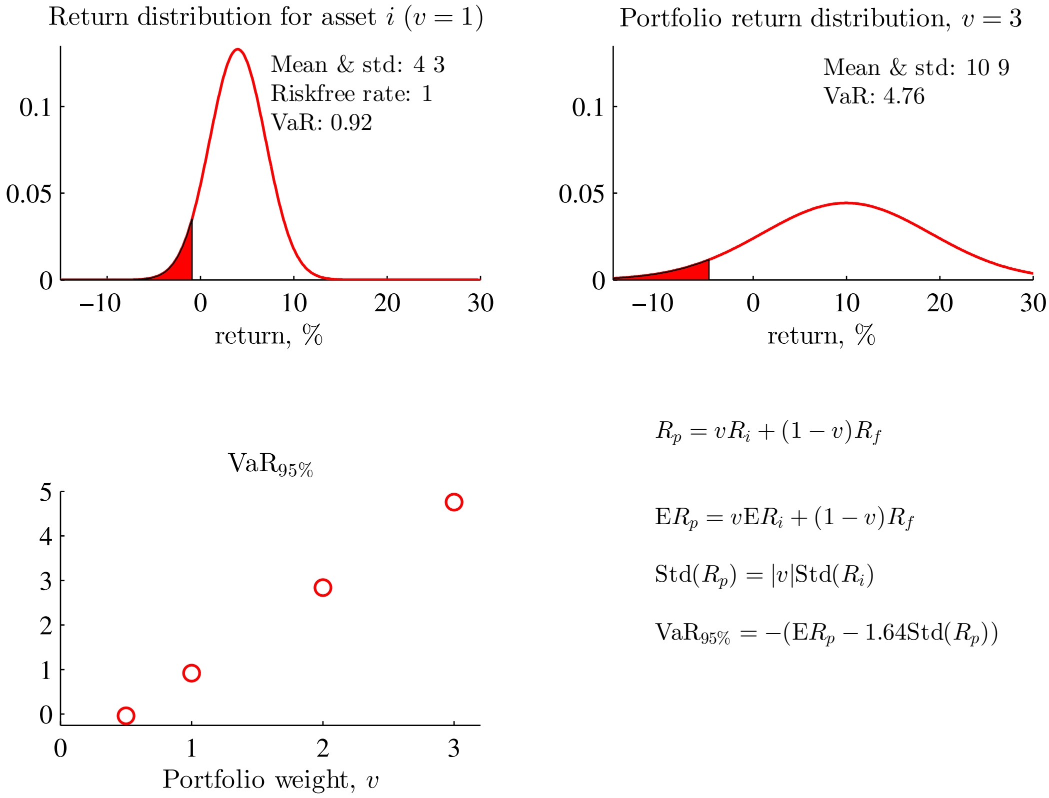


Figure 4.9: The effect of leverage on the portfolio return distribution and VaR

For a normally distributed return $R \sim N(\mu, \sigma^2)$ we have

$$\text{ES}_\alpha = -\mu + \sigma \frac{\phi(c_{1-\alpha})}{1-\alpha}, \tag{4.17}$$

where $\phi()$ is the pdf or a $N(0,1)$ variable and where $c_{1-\alpha}$ is the $1-\alpha$ quantile of a N(0,1) distribution (for instance, -1.64 for $1-\alpha = 0.05$).

Proof. (of (4.17)) If $x \sim N(\mu, \sigma^2)$, then $\text{E}(x|x \leq b) = \mu - \sigma\phi(b_0)/\Phi(b_0)$ where $b_0 = (b-\mu)/\sigma$ and where $\phi()$ and $\Phi()$ are the pdf and cdf of a $N(0,1)$ variable respectively. To apply this, use $b = -\text{VaR}_\alpha$ so $b_0 = c_{1-\alpha}$. Clearly, $\Phi(c_{1-\alpha}) = 1-\alpha$ (by definition of the $1-\alpha$ quantile). Multiply by -1. $\blacksquare$

Example 4.7 *(ES) If $\mu = 8\%$ and $\sigma = 16\%$, the 95% expected shortfall is $ES_{95\%} = -0.08 + 0.16\phi(-1.64)/0.05 \approx 0.25$ and the 97.5% expected shortfall is $ES_{97.5\%} = -0.08 + 0.16\phi(-1.96)/0.025 \approx 0.29$.*

Notice that the expected shortfall for a normally distributed return (4.17) is a strictly increasing function of the standard deviation (and the variance). Minimizing the expected shortfall at a given mean return therefore gives the same solution (portfolio weights) as minimizing the variance at the same given mean return. In other cases, the portfolio choice will be different (and perhaps complicated to perform).

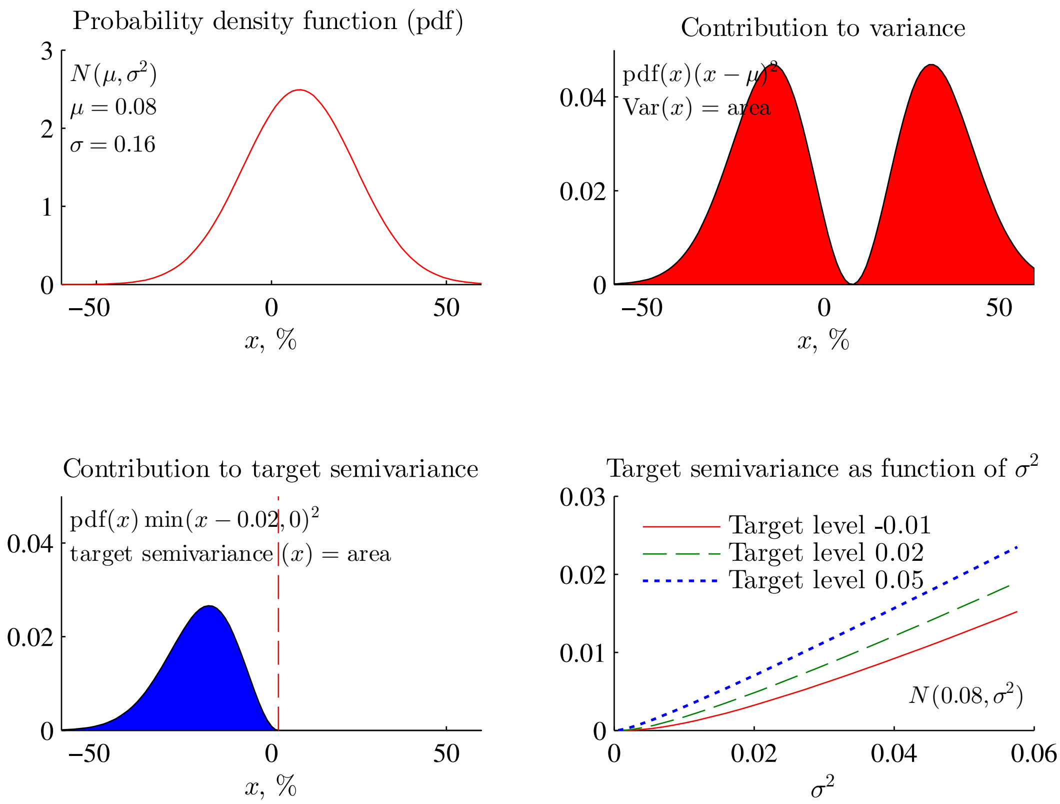


Figure 4.10: Target semivariance as a function of mean and standard deviation for a N(μ,σ^2) variable

4.2.7 Target Semivariance (Lower Partial 2nd Moment) and Max Drawdown

Reference: Bawa and Lindenberg (1977) and Nantell and Price (1979)

Using the variance (or standard deviation) as a measure of portfolio risk (as a mean-variance investor does) fails to distinguish between the downside and upside. As an alternative, one could consider using a target semivariance (lower partial 2nd moment) instead.

It is defined as

$$\lambda_p(h) = \mathrm{E}[\min(R_p - h, 0)^2], \tag{4.18}$$

where h is a "target level" chosen by the investor. In the subsequent analysis it will be set equal to the riskfree rate. (It can clearly also be written $\lambda_p(h) = \int_{-\infty}^{h} (R_p - h)^2 f(R_p) dR_p$, where $f()$ is the pdf of the portfolio return.) The square root of $\lambda(\mathrm{E}\, R_p)$ is called the semi-standard deviation.

In comparison with a variance

$$\sigma_p^2 = \mathrm{E}(R_p - \mathrm{E}\, R_p)^2, \tag{4.19}$$

the target semivariance differs on two accounts: *(i)* it uses the target level h as a reference point instead of the mean $\mathrm{E}\, R_p$: and *(ii)* only negative deviations from the reference point are given any weight. See Figure 4.10 for an illustration (based on a normally distributed variable).

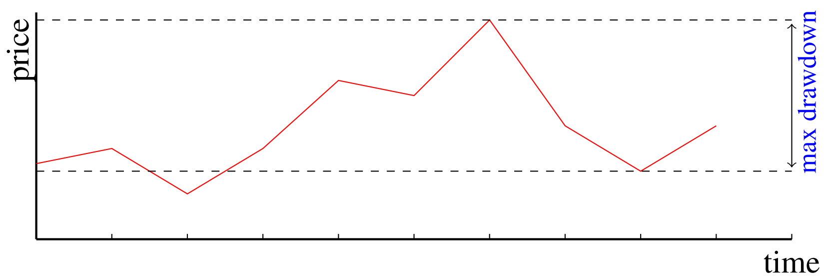


Figure 4.11: Max drawdown

For a normally distributed variable, the target semivariance $\lambda_p(h)$ is increasing in the standard deviation (for a given mean)—see Remark 4.8. See also Figure 4.10 for an illustration.

An alternative measure is the (percentage) *maximum drawdown* over a given horizon, for instance, 5 years, say. This is the largest loss from peak to bottom within the given horizon–see Figure 4.11. This is a useful measure when the investor do not know exactly when he/she has to exit the investment—since it indicates the worst (peak to bottom) outcome over the sample.

See Figures 4.12–4.13 for an illustration of max drawdown.

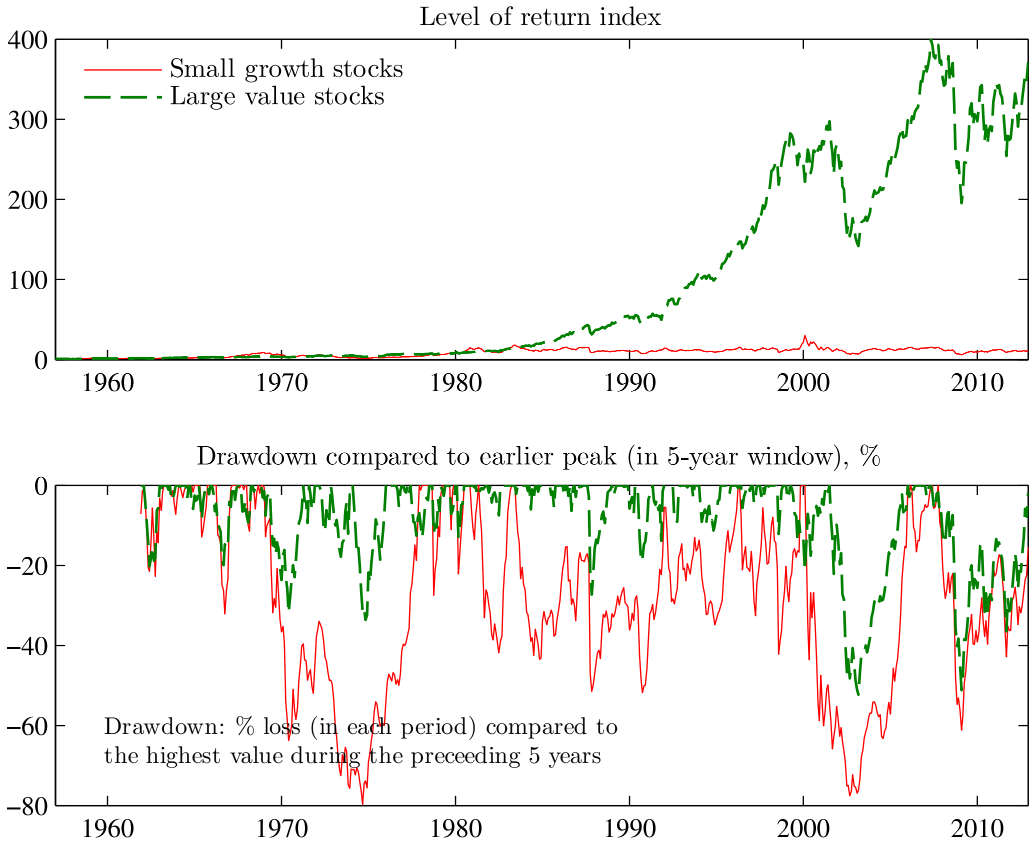


Figure 4.12: Drawdown

Remark 4.8 *(Target semivariance calculation for normally distributed variable*) For an $N(\mu, \sigma^2)$ variable, target semivariance around the target level h is*

$$\lambda_p(h) = \sigma^2 a\phi(a) + \sigma^2(a^2 + 1)\Phi(a), \text{ where } a = (h - \mu)/\sigma,$$

where $\phi()$ and $\Phi()$ are the pdf and cdf of a $N(0, 1)$ variable respectively. Notice that $\lambda_p(h) = \sigma^2/2$ for $h = \mu$. See Figure 4.10 for a numerical illustration. It is straightforward (but a bit tedious) to show that

$$\frac{\partial \lambda_p(h)}{\partial \sigma} = 2\sigma\Phi(a),$$

so the target semivariance is a strictly increasing function of the standard deviation.

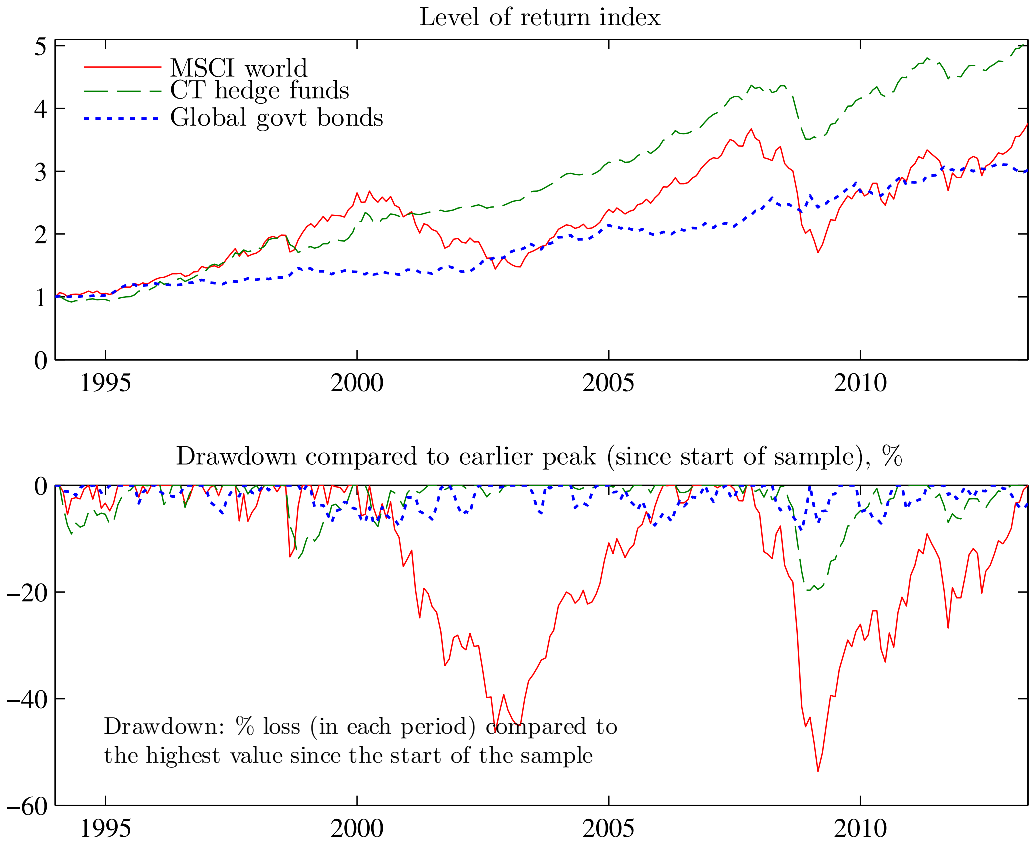


Figure 4.13: Drawdown

See Table 4.2 for an empirical comparison of the different risk measures.

	Std	VaR (95%)	ES (95%)	SemiStd	Drawdown
Std	1.00	0.94	0.98	0.97	0.68
VaR (95%)	0.94	1.00	0.94	0.95	0.72
ES (95%)	0.98	0.94	1.00	0.98	0.67
SemiStd	0.97	0.95	0.98	1.00	0.68
Drawdown	0.68	0.72	0.67	0.68	1.00

Table 4.2: Correlation of rank of risk measures across the 25 FF portfolios (%), US data 1957:1-2012:12.

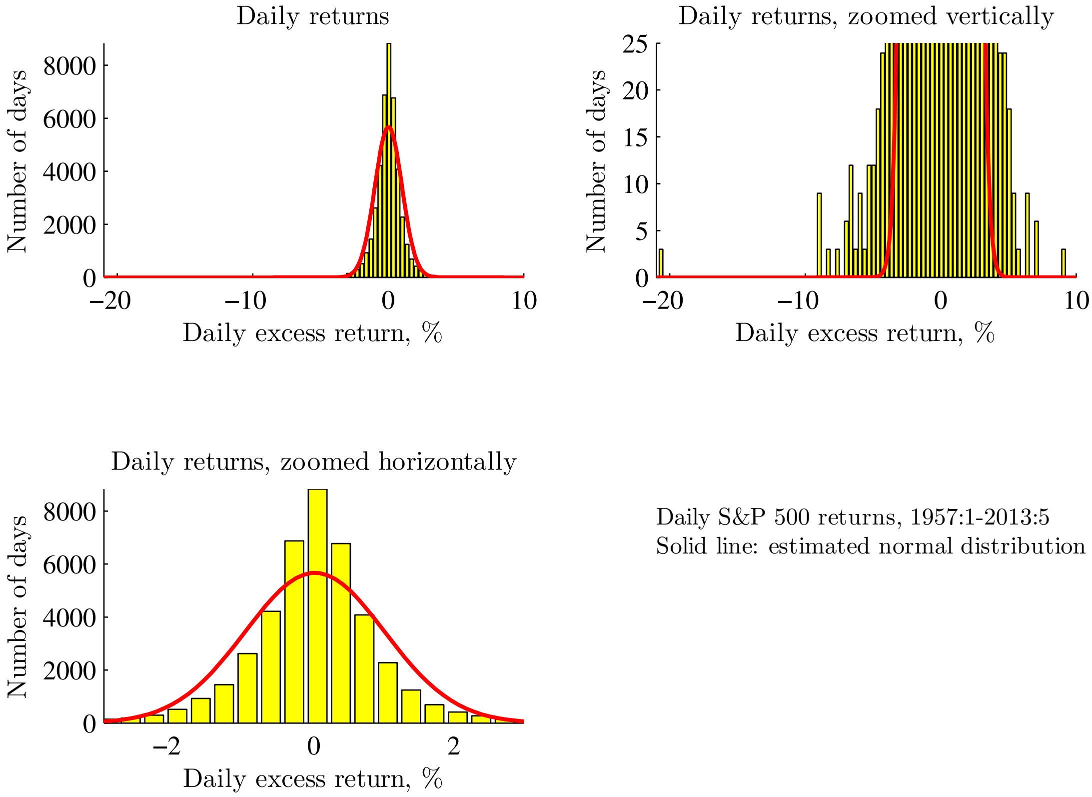


Figure 4.14: Distribution of daily S&P returns

4.3 Empirical Return Distributions

Are returns normally distributed? Mostly not, but it depends on the asset type and on the data frequency. Options returns typically have very non-normal distributions (in particular, since the return is -100% on many expiration days). Stock returns are typically distinctly non-linear at short horizons, but can look somewhat normal at longer horizons.

To assess the normality of returns, the usual econometric techniques (Bera–Jarque and Kolmogorov-Smirnov tests) are useful, but a visual inspection of the histogram and a QQ-plot also give useful clues. See Figures 4.14–4.16 for illustrations.

Remark 4.9 *(Reading a QQ plot) A QQ plot is a way to assess if the empirical distribution conforms reasonably well to a prespecified theoretical distribution, for instance, a normal distribution where the mean and variance have been estimated from the data. Each point in the QQ plot shows a specific percentile (quantile) according to the empiri-*

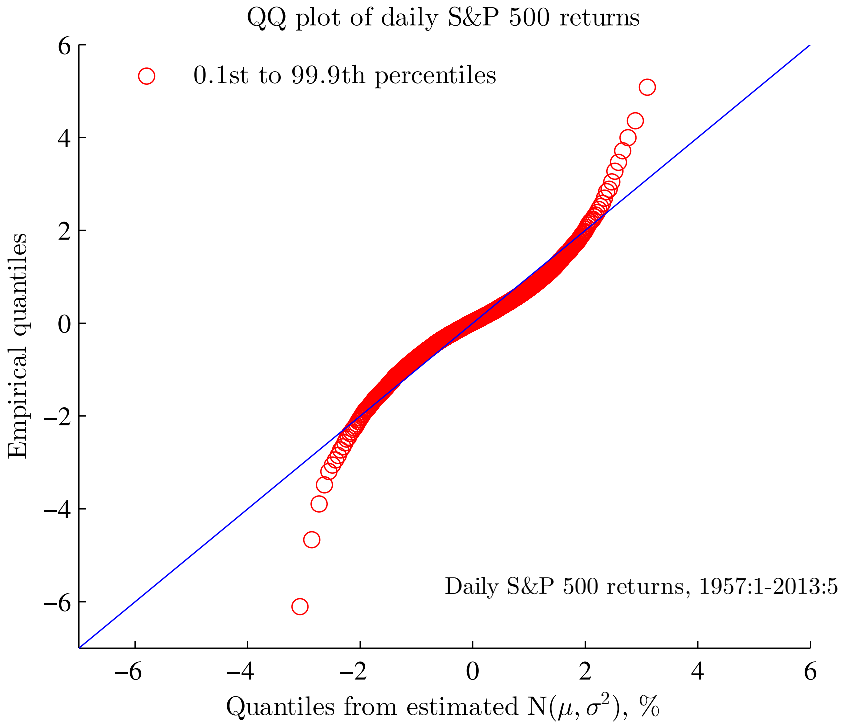


Figure 4.15: Quantiles of daily S&P returns

cal as well as according to the theoretical distribution. For instance, if the 2th percentile (0.02 percentile) is at -10 in the empirical distribution, but at only -3 in the theoretical distribution, then this indicates that the two distributions have fairly different left tails.

There is one caveat to this way of studying data: it only provides evidence on the unconditional distribution. For instance, nothing rules out the possibility that we could estimate a model for time-varying volatility (for instance, a GARCH model) of the returns and thus generate a description for how the VaR changes over time. However, data with time varying volatility will typically not have an unconditional normal distribution.

Bibliography

Alexander, C., 2008, *Market Risk Analysis: Value at Risk Models*, Wiley.

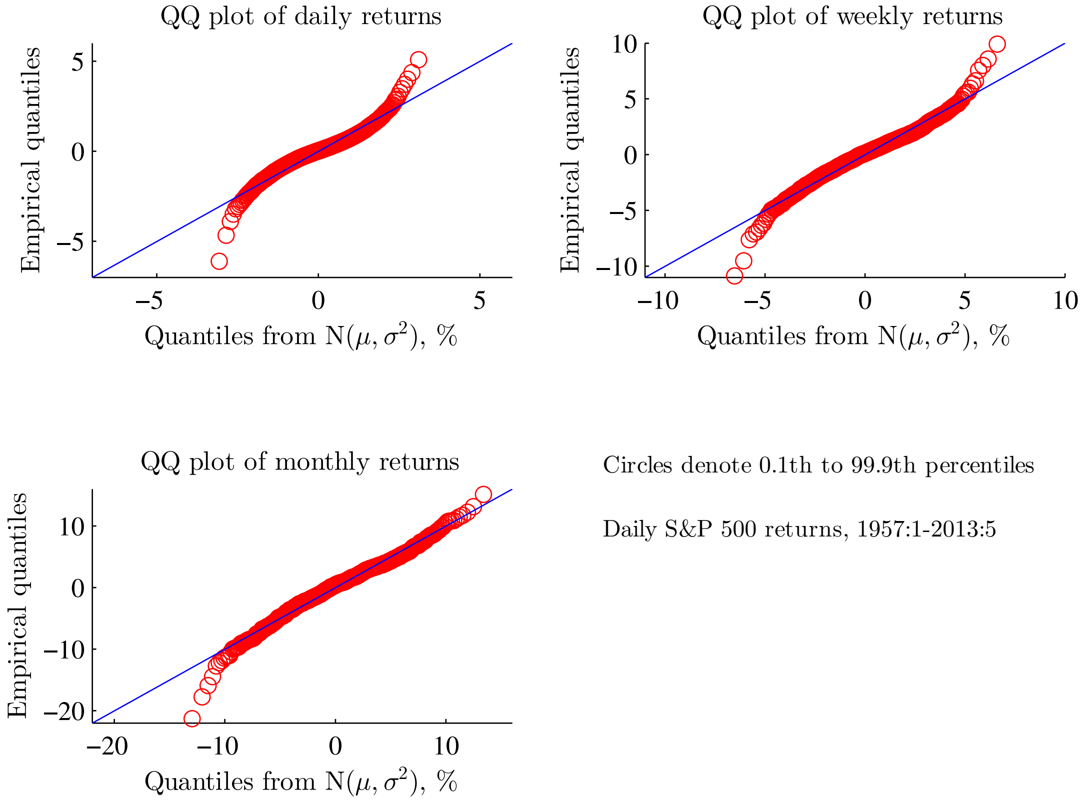


Figure 4.16: Distribution of S&P returns (different horizons)

Bawa, V. S., and E. B. Lindenberg, 1977, "Capital market equilibrium in a mean-lower partial moment framework," *Journal of Financial Economics*, 5, 189–200.

Fabozzi, F. J., S. M. Focardi, and P. N. Kolm, 2006, *Financial modeling of the equity market*, Wiley Finance.

Hull, J. C., 2006, *Options, futures, and other derivatives*, Prentice-Hall, Upper Saddle River, NJ, 6th edn.

McDonald, R. L., 2006, *Derivatives markets*, Addison-Wesley, 2nd edn.

McNeil, A. J., R. Frey, and P. Embrechts, 2005, *Quantitative risk management*, Princeton University Press.

Nantell, T. J., and B. Price, 1979, "An analytical comparison of variance and semivariance capital market theories," *Journal of Financial and Quantitative Analysis*, 14, 221–242.

5 CAPM

Reference: Elton, Gruber, Brown, and Goetzmann (2010) 10 and 13

Additional references: Danthine and Donaldson (2002) 6

More advanced material is denoted by a star (*). It is not required reading.

5.1 Portfolio Choice with Mean-Variance Utility

It is well known that mean-variance preferences (and several other cases) imply that the optimal portfolio is a mix of the riskfree asset and the tangency portfolio (a portfolio of risky assets only) that is located at the point where the ray from the riskfree rate is tangent to the mean-variance frontier of risky assets only. See Figure 5.1 for an example. The purpose of this section is to derive a formula for the tangency portfolio.

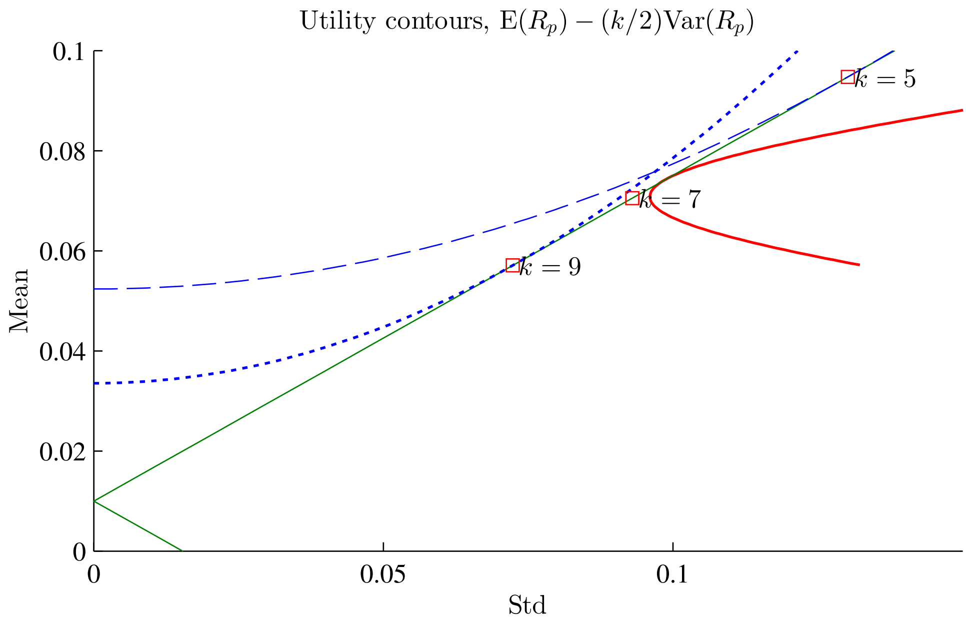


Figure 5.1: Iso-utility curves, mean-variance utility

5.1.1 A Risky Asset and a Riskfree Asset (recap)

Suppose there are one risky asset and a riskfree asset. An investor with initial wealth equal (to simplify the notation) to unity chooses the portfolio weight v (of the risky asset) to maximize

$$\mathrm{E}\, U(R_p) = \mathrm{E}\, R_p - \frac{k}{2}\operatorname{Var}(R_p), \text{ where} \tag{5.1}$$

$$R_p = v R_1^e + R_f. \tag{5.2}$$

We have already demonstrated that the optimal portfolio weight of the risky asset is

$$v = \frac{1}{k}\frac{\mu_1^e}{\sigma_{11}}. \tag{5.3}$$

Clearly, the weight on the risky asset is increasing in the expected excess return of the risky asset, but decreasing in the risk aversion and variance.

We have also show that the optimal solution implies that

$$\frac{\mathrm{E}\, R_p^e}{\operatorname{Var}(R_p)} = k, \tag{5.4}$$

where R_p is the portfolio return (5.2) obtained by using the optimal v (from (5.3)). It shows that an investor with a high risk aversion (k) will choose a portfolio with a high return compared to the volatility.

Figures 5.2–5.3 illustrate the effect on the portfolio return distribution.

5.1.2 Two Risky Assets and a Riskfree Asset

With two risky assets, we can analyze the effect of correlations of returns.

We now go through the same steps for the case with two risky assets and a riskfree asset. An investor (with initial wealth equal to unity) chooses the portfolio weights (v_i) to maximize

$$\mathrm{E}\, U(R_p) = \mathrm{E}\, R_p - \frac{k}{2}\operatorname{Var}(R_p), \text{ where} \tag{5.5}$$

$$\begin{aligned} R_p &= v_1 R_1 + v_2 R_2 + (1 - v_1 - v_2) R_f \\ &= v_1 R_1^e + v_2 R_2^e + R_f. \end{aligned} \tag{5.6}$$

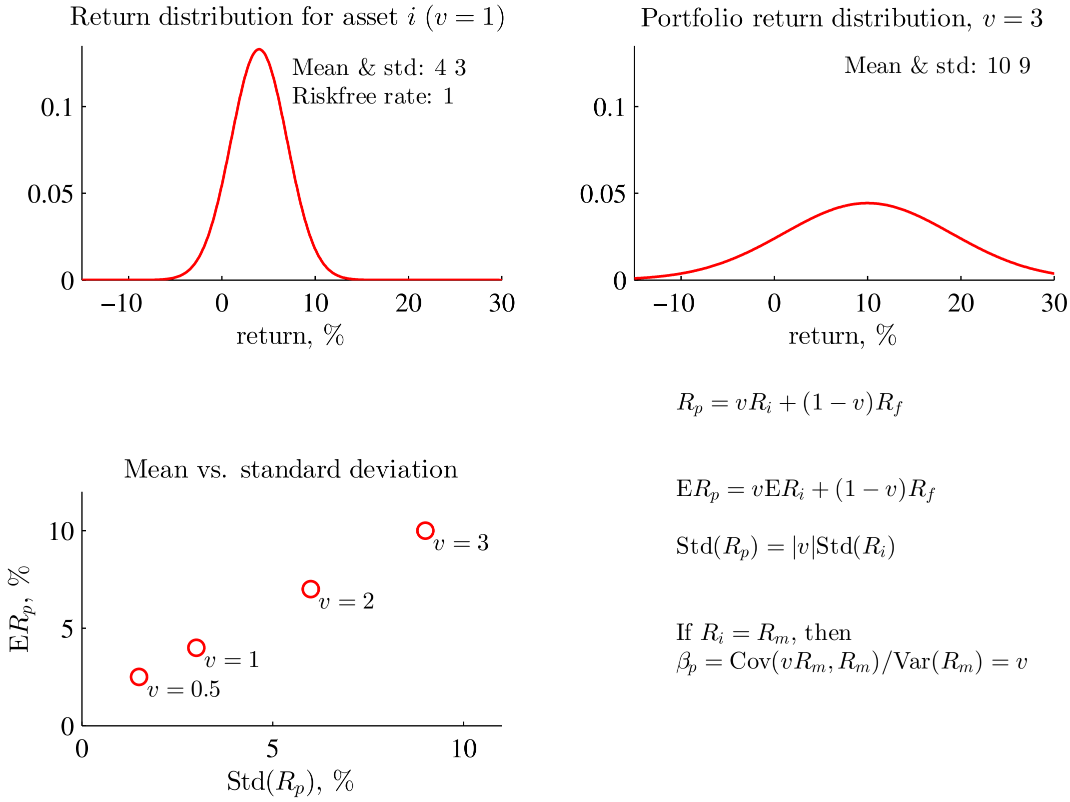


Figure 5.2: The effect of leverage on the portfolio return distribution

Combining gives

$$\begin{aligned} \mathrm{E}\,U(R_p) &= \mathrm{E}(v_1 R_1^e + v_2 R_2^e + R_f) - \frac{k}{2}\,\mathrm{Var}(v_1 R_1^e + v_2 R_2^e + R_f) \\ &= v_1 \mu_1^e + v_2 \mu_2^e + R_f - \frac{k}{2}\left(v_1^2 \sigma_{11} + v_2^2 \sigma_{22} + 2 v_1 v_2 \sigma_{12}\right), \end{aligned} \tag{5.7}$$

where σ_{12} denotes the covariance of asset 1 and 2.

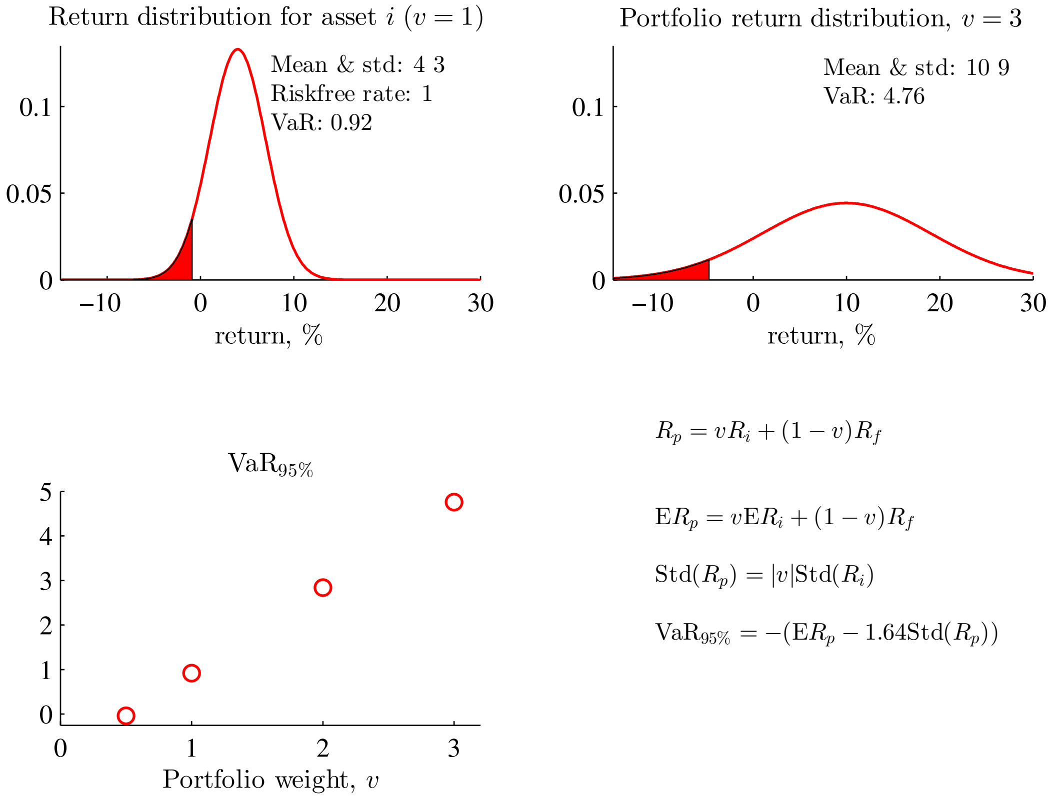


Figure 5.3: The effect of leverage on the portfolio return distribution and VaR

The first order conditions (for v_1 and v_2) are that the partial derivatives equal zero

$$0 = \partial \mathrm{E}\, U(R_p)/\partial v_1 = \mu_1^e - \frac{k}{2}\left(2v_1\sigma_{11} + 2v_2\sigma_{12}\right) \tag{5.8}$$

$$0 = \partial \mathrm{E}\, U(R_p)/\partial v_2 = \mu_2^e - \frac{k}{2}\left(2v_2\sigma_{22} + 2v_1\sigma_{12}\right), \text{ or} \tag{5.9}$$

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \mu_1^e \\ \mu_2^e \end{bmatrix} - k \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \tag{5.10}$$

$$\mathbf{0}_{2\times 1} = \mu^e - k\Sigma v. \tag{5.11}$$

We can solve this linear system of equations as

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \frac{1}{k}\frac{1}{\sigma_{11}\sigma_{22}-\sigma_{12}^2}\begin{bmatrix} \sigma_{22}\mu_1^e - \sigma_{12}\mu_2^e \\ -\sigma_{12}\mu_1^e + \sigma_{11}\mu_2^e \end{bmatrix} \tag{5.12}$$

$$= \frac{1}{k}\frac{1}{\sigma_{11}\sigma_{22}-\sigma_{12}^2}\begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}\begin{bmatrix} \mu_1^e \\ \mu_2^e \end{bmatrix} \tag{5.13}$$

$$= \frac{1}{k}\Sigma^{-1}\mu^e, \tag{5.14}$$

where Σ is the covariance matrix and μ^e the vector of excess returns.

Notice that the denominator $(\sigma_{11}\sigma_{22}-\sigma_{12}^2)$ is positive—since correlations are between -1 and 1. Since $k > 0$, we have

$$v_1 > 0 \text{ if } \sigma_{22}\mu_1^e > \sigma_{12}\mu_2^e. \tag{5.15}$$

Use the fact that $\sigma_{12} = \rho\sigma_1\sigma_2$ where ρ is the correlation coefficient to rewrite as

$$v_1 > 0 \text{ if } \mu_1^e/\sigma_1 > \rho\mu_2^e/\sigma_2, \text{ and} \tag{5.16}$$

$$v_2 > 0 \text{ if } \mu_2^e/\sigma_2 > \rho\mu_1^e/\sigma_1. \tag{5.17}$$

This provides a simple way to assess if an asset should be held (in positive amounts): if its Sharpe ratio exceeds the correlation times the Sharpe ratio of the other asset. For instance, both portfolio weights are positive if the correlation is zero and both excess returns are positive.

For some value of the risk aversion k, the portfolio weights in (5.14) sum to one, so there is no investment in the riskfree asset. This holds for

$$k_T = \mathbf{1}'\Sigma^{-1}\mu^e, \tag{5.18}$$

where $\mathbf{1}$ is a vector of ones (clearly, k_T is a scalar). In this case, (5.12)–(5.14) become

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \sigma_{22}\mu_1^e - \sigma_{12}\mu_2^e \\ -\sigma_{12}\mu_1^e + \sigma_{11}\mu_2^e \end{bmatrix}\frac{1}{\sigma_{22}\mu_1^e + \sigma_{11}\mu_2^e - (\mu_2^e + \mu_1^e)\sigma_{12}} \tag{5.19}$$

$$= \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}\begin{bmatrix} \mu_1^e \\ \mu_2^e \end{bmatrix}\frac{1}{\sigma_{22}\mu_1^e + \sigma_{11}\mu_2^e - (\mu_2^e + \mu_1^e)\sigma_{12}} \tag{5.20}$$

$$= \Sigma^{-1}\mu^e/\mathbf{1}'\Sigma^{-1}\mu^e, \tag{5.21}$$

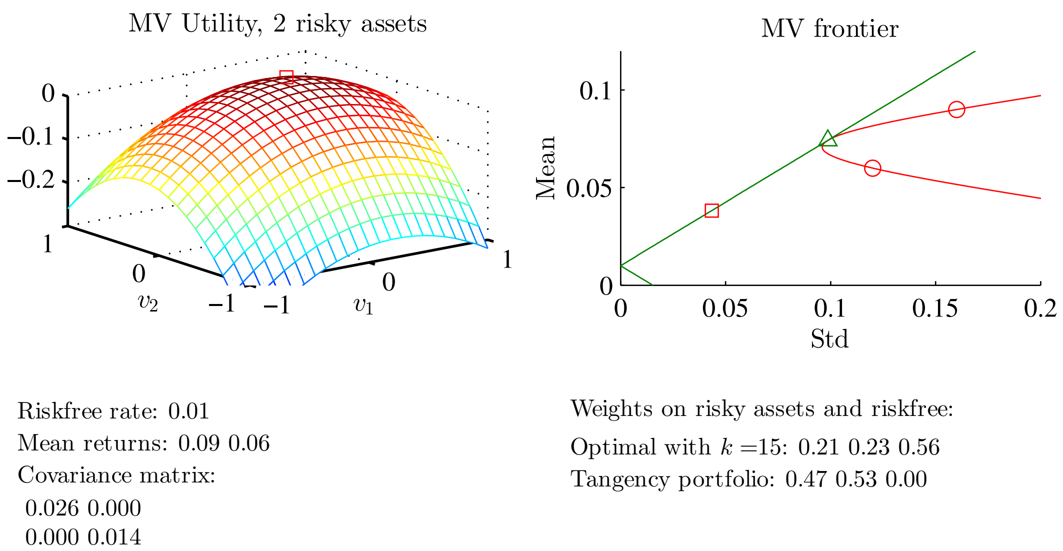


Figure 5.4: Choice of portfolios weights

This is actually the *tangency portfolio* from mean-variance analysis (where the ray from R_f in the $\mathrm{Std}(R_p) \times \mathrm{E}\, R_p$ space is tangent to the minimum-variance set). It has the highest Sharpe ratio, $\mathrm{E}\, R_p^e / \mathrm{Std}(R_p)$, of all portfolios on the minimum-variance set. See Figure 5.4 for an illustration.

Note that all investors (different k, but same expectations) hold a mix of this portfolio and the riskfree asset. To see that, notice that (5.14) can be written

$$v = \frac{k_T}{k} w, \tag{5.22}$$

where k_T is defined in (5.18) and where w is the vector of weights in the tangency portfolio (from (5.21)). Since the first term on the right hand side (k_T/k) is a scalar, this shows that every investor holds a scaled version of the tangency portfolio. The balance $(1 - \mathbf{1}'v)$ is made up by a position in the riskfree asset. This *two-fund separation theorem* is very useful. This means that all investors are on the MV frontier (including a riskfree asset), also called the capital market line (CML). To see this, notice that (a) when $k = k_T$ then the investor is at the tangency portfolio; (b) when $k = \infty$ then the investor only invests in the riskfree asset. For all intermediate values of k the investor is on the straight line between the riskfree asset and the tangency portfolio (or beyond it if $k < k_T$).

Consider the simple case when the assets are uncorrelated ($\sigma_{12} = 0$), then the tangency portfolio (5.19) becomes

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \sigma_{22}\mu_1^e \\ \sigma_{11}\mu_2^e \end{bmatrix} \frac{1}{\sigma_{22}\mu_1^e + \sigma_{11}\mu_2^e}. \tag{5.23}$$

Results: *(i)* if both excess returns are positive, then the weight on asset 1 increases if μ_1^e increases or σ_{11} decreases; *(ii)* both weights are positive if the excess returns are. Both results are quite intuitive since the investor likes high expected returns, but dislikes variance.

Example 5.1 *(Tangency portfolio, numerical) When* $(\mu_1^e, \mu_2^e) = (0.08, 0.05)$*, the correlation is zero, and* $(\sigma_{11}, \sigma_{22}) = (0.16^2, 0.12^2)$*, then (5.23) gives*

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0.47 \\ 0.53 \end{bmatrix}.$$

When μ_1^e *increases from* 0.08 *to* 0.12*, then we get*

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0.57 \\ 0.43 \end{bmatrix}.$$

Now, consider another simple case, where both variances are the same, but the correlation is non-zero ($\sigma_{11} = \sigma_{22} = 1$ as a normalization, $\sigma_{12} = \rho$). Then (5.19) becomes

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \mu_1^e - \rho\mu_2^e \\ \mu_2^e - \rho\mu_1^e \end{bmatrix} \frac{1}{(\mu_1^e + \mu_2^e)(1-\rho)}. \tag{5.24}$$

Results: *(i)* both weights are positive if the returns are negatively correlated ($\rho < 0$) and both excess returns are positive; *(ii)* $w_2 < 0$ if $\rho > 0$ and μ_1^e is considerably higher than μ_2^e (so $\mu_2^e < \rho\mu_1^e$). The intuition for the first result is that a negative correlation means that the assets "hedge" each other (even better than diversification), so the investor would like to hold both of them to reduce the overall risk. (Unfortunately, most assets tend to be positively correlated.) The intuition for the second result is that a positive correlation reduces the gain from holding both assets (they don't hedge each other, and there is relatively little diversification to be gained if the correlation is high). On top of this, asset 1 gives a higher expected return, so it is optimal to sell asset 2 short (essentially a risky "loan" which allows the investor to buy more of asset 1).

Example 5.2 *(Tangency portfolio, numerical) When* $(\mu_1^e, \mu_2^e) = (0.08, 0.05)$*, and* $\rho = -0.8$ *we get*

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0.51 \\ 0.49 \end{bmatrix}.$$

If, instead, $\rho = 0.8$*, then we get*

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 1.54 \\ -0.54 \end{bmatrix}.$$

5.1.3 N Risky Assets and a Riskfree Asset

In the general case with N risky assets and a riskfree asset, the portfolio weights of the risky assets are

$$v = \frac{1}{k}\Sigma^{-1}\mu^e, \tag{5.25}$$

while the weight on the riskfree asset is $1 - \mathbf{1}'v$. The weights of the tangency portfolio (where $\mathbf{1}'v = 1$) are therefore

$$w_T = \Sigma^{-1}\mu^e / \mathbf{1}'\Sigma^{-1}\mu^e. \tag{5.26}$$

As before, we can write the portfolio weights v as scaled versions of the tangency portfolio

$$v = \frac{k_T}{k} w_T, \tag{5.27}$$

where $k_T = \mathbf{1}'\Sigma^{-1}\mu^e$ (a scalar) is the risk aversion that would make the investor hold only risky assets (no riskfree).

Proof. (of (5.25)–(5.26)) The portfolio has the return $R_p = v'R + (1 - \mathbf{1}'v)R_f = v'(R - R_f) + R_f$. The mean and variance are

$$\mathrm{E}\, R_p = v'\mu^e + R_f \text{ and } \mathrm{Var}(R_p) = v'\Sigma v.$$

The optimization problem is

$$\max_v v'\mu^e + R_f - \frac{k}{2} v'\Sigma v,$$

with first order conditions (see Appendix for matrix calculus)

$$\mathbf{0}_{N\times 1} = \mu^e - k\Sigma v,$$

which gives (5.25).

To prove (5.26), notice that to have $\mathbf{1}'v = 1$, (5.25) says that $\mathbf{1}'\Sigma^{-1}\mu^e = k_T$ must hold. Combine with (5.25) to get (5.26). ■

As in the case with only one risky asset, the optimal portfolio (v) has

$$\begin{aligned} \frac{\mathrm{E}\, R_p^e}{\mathrm{Var}(R_p)} &= k, \text{ and} \\ SR_p &= \sqrt{\mu^{e\prime}\Sigma^{-1}\mu^e}, \end{aligned} \tag{5.28}$$

which SR_p is the Sharpe ratio of the portfolio. The first line says that higher risk aversion tilts the portfolio away from a high variance—and the second line says that all investors (irrespective of their risk aversions) have the same Sharpe ratios. This is clearly the same as saying that they all mix the tangency portfolio with the risk free asset (depending on their risk aversion)—they are all on the Capital Market Line (see Figure 5.11). Clearly, with $k = \infty$, the portfolio has a zero variance, so the expected excess return is zero. With lower risk aversion, the portfolio shifts along the CLM towards higher variance (and expected return).

Proof. (of (5.28)) Use the portfolio weights in (5.25) to write

$$\begin{aligned} \frac{\mathrm{E}\, R_p^e}{\mathrm{Var}(R_p)} &= \frac{\left(\frac{1}{k}\Sigma^{-1}\mu^e\right)'\mu^e}{\left(\frac{1}{k}\Sigma^{-1}\mu^e\right)'\Sigma\left(\frac{1}{k}\Sigma^{-1}\mu^e\right)} = k. \\ &= k\frac{\left(\Sigma^{-1}\mu^e\right)'\mu^e}{\left(\Sigma^{-1}\mu^e\right)'\mu^e} \end{aligned}$$

Multiply by $\mathrm{Std}(R_p)$ to get the Sharpe ratio of the portfolio

$$\begin{aligned} SR_p &= k\,\mathrm{Std}(R_p) \\ &= k\sqrt{\left(\frac{1}{k}\Sigma^{-1}\mu^e\right)'\Sigma\left(\frac{1}{k}\Sigma^{-1}\mu^e\right)} \\ &= \sqrt{\mu^{e\prime}\Sigma^{-1}\mu^e}. \end{aligned}$$

■

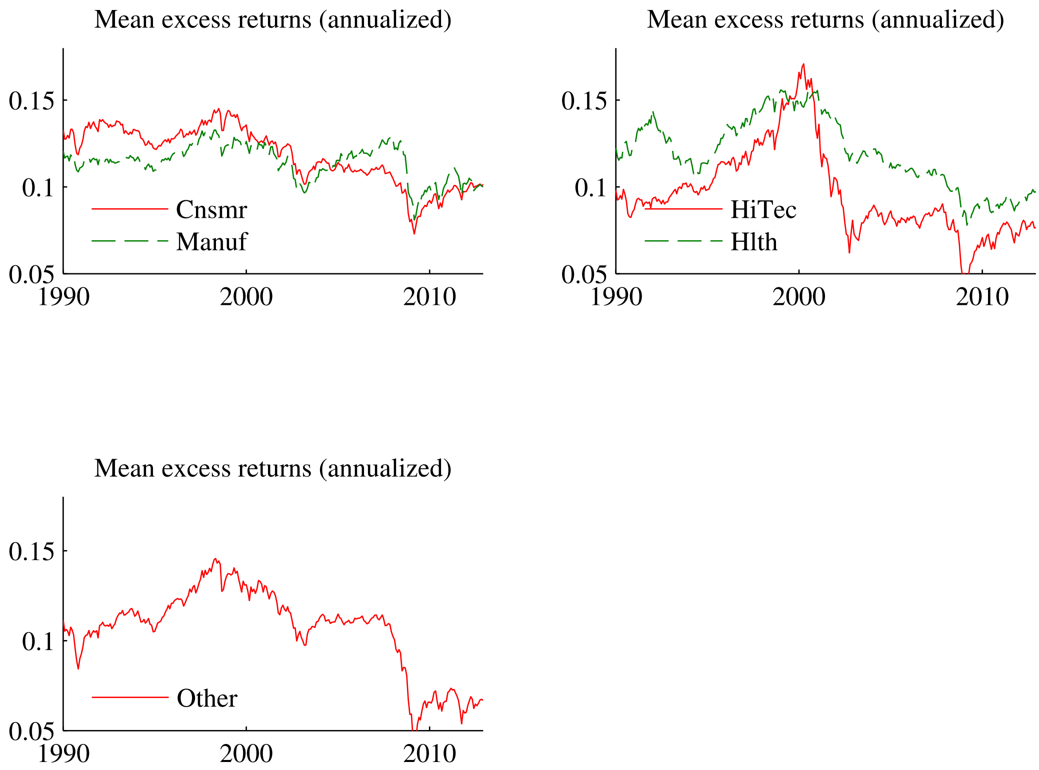


Figure 5.5: Dynamicically updated estimates, 5 U.S. industries

Remark 5.3 *(Properties of tangency portfolio) The expected excess return and the variance of the tangency portfolio are* $\mu_T^e = \mu^{e\prime}\Sigma^{-1}\mu^e / \mathbf{1}'\Sigma^{-1}\mu^e$ *and* $\operatorname{Var}(R_T^e) = \mu^{e\prime}\Sigma^{-1}\mu^e / \left(\mathbf{1}'\Sigma^{-1}\mu^e\right)^2$*. The square of the Sharpe ratio is therefore* $\left(\mu_T^e/\sigma_T\right)^2 = \mu^{e\prime}\Sigma^{-1}\mu^e$*.*

5.1.4 Historical Estimates of the Average Returns and the Covariance Matrix

Figures 5.5–5.6 illustrate mean returns and standard deviations, estimated by exponentially moving averages (as by RiskMetrics). Figures 5.7–5.8 show how the optimal portfolio weights (based on mean-variance preferences). It is clear that the portfolio weights change very dramatically—perhaps too much to be realistic. It is also clear that the changes in estimated average returns cause more dramatic movements in the portfolio weights than the changes in the estimated covariance matrix.

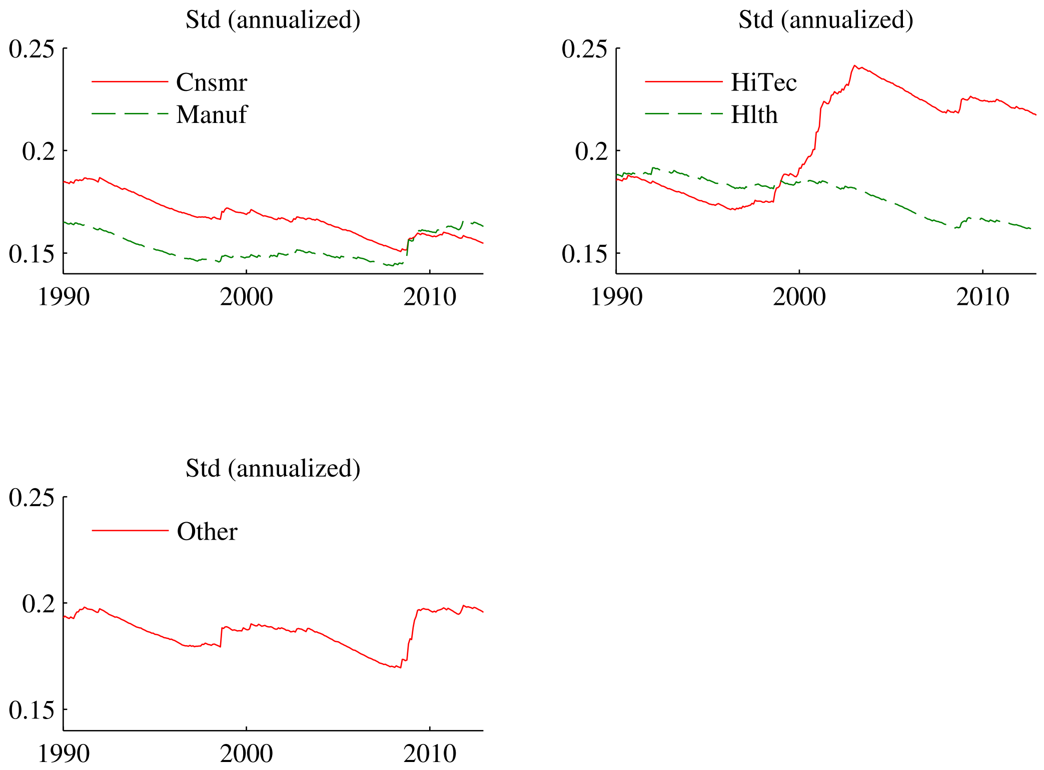


Figure 5.6: Dynamicically updated estimates, 5 U.S. industries

5.1.5 A Risky Asset and a Riskfree Asset Revisited

Once we have the tangency portfolio (with weights w as in (5.26)), we can actually use that as *the* risky asset in the case with only one risky asset (and a riskfree). That is, we can treat $w'R^e$ as R_1^e in (5.2). After all, the portfolio choice is really about mixing the tangency portfolio with the riskfree asset.

The result is that the weight on the tangency portfolio is (a scalar)

$$v^* = \frac{1}{k}\mathbf{1}'\Sigma^{-1}\mu^e, \tag{5.29}$$

and $1 - v^*$ on the riskfree asset.

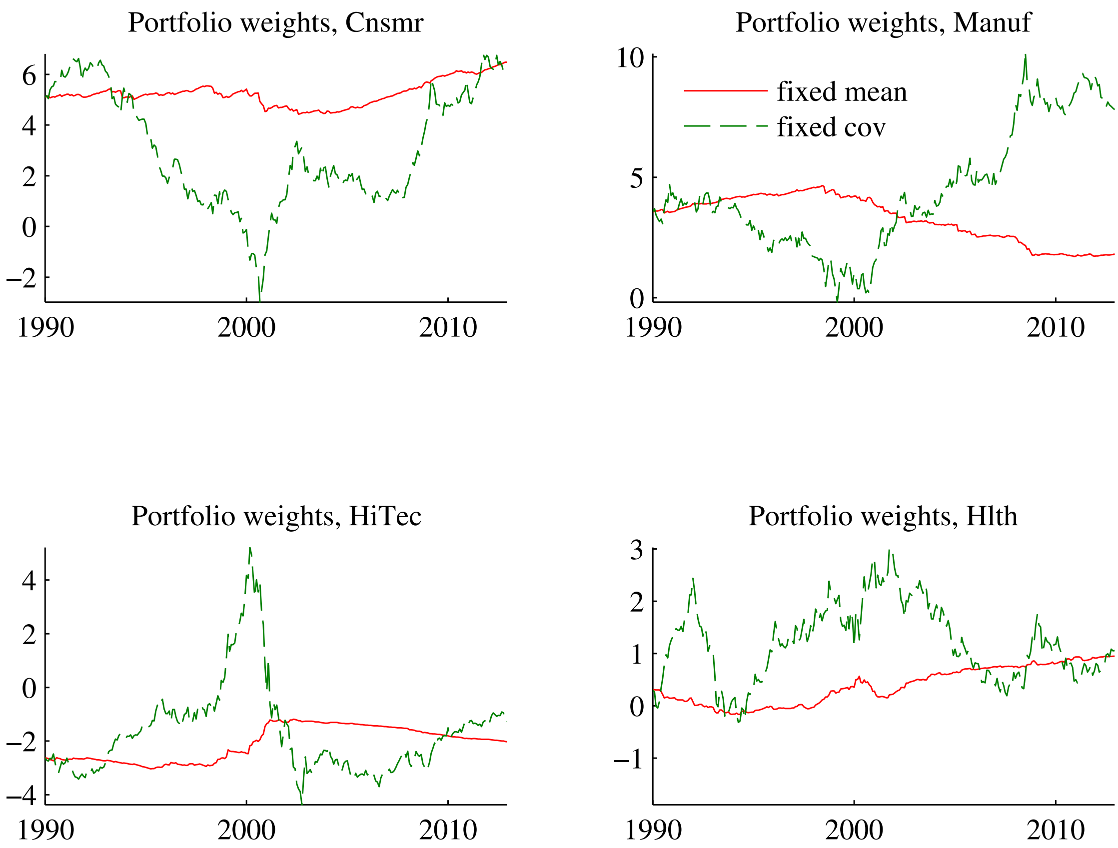


Figure 5.7: Dynamicically updated portfolio weights, T-bill and 5 U.S. industries

Proof. (of (5.29)) From (5.25)–(5.26) we directly get

$$v = \underbrace{\frac{1}{k}\mathbf{1}'\Sigma^{-1}\mu^e}_{v^*} w,$$

which is just v^* in (5.29) times the tangency portfolio w from (5.26). To see that this fits with (5.3) when $w'R^e$ is substituted for R_1^e, notice that

$$\frac{\mathrm{E}\, w'R^e}{\mathrm{Var}(w'R)} = \mathbf{1}'\Sigma^{-1}\mu^e,$$

so (5.3) could be written just like (5.29). ■

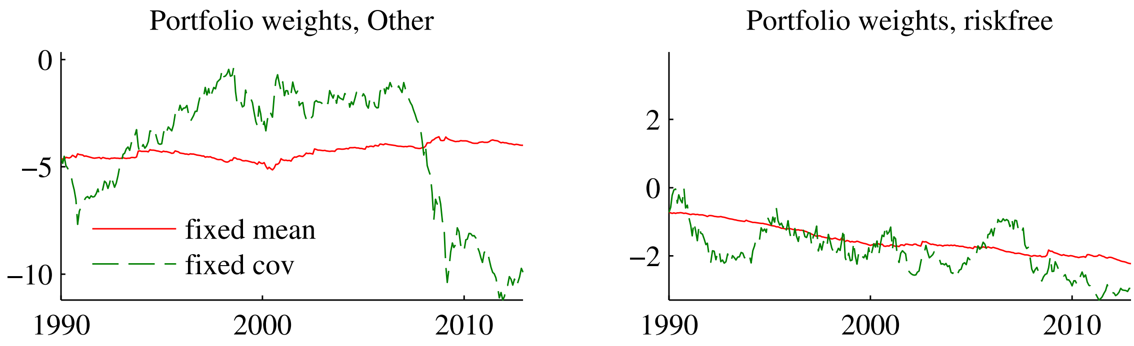


Figure 5.8: Dynamicically updated portfolio weights, T-bill and 5 U.S. industries

5.1.6 Portfolio Choice with Short Sale Constraints

The previous analysis assumes that there are no restrictions on the portfolio weights. However, many investors (for instance, mutual funds) cannot have short positions. In this case, the objective function is still (5.5), but with the additional restriction

$$0 \leq v_i \leq 1. \tag{5.30}$$

See Figures 5.9–5.10 for an illustration.

5.2 Beta Representation of Expected Returns

For any portfolio, the expected excess return ($\mathrm{E}\, R_p^e$) is linearly related to the expected excess return on the tangency portfolio (μ_T^e) according to

$$\mathrm{E}\, R_p^e = \beta_p \mu_T^e \text{, where } \beta_p = \frac{\mathrm{Cov}\left(R_p, R_T\right)}{\mathrm{Var}\left(R_T\right)}. \tag{5.31}$$

This result follows directly from manipulating the definition of the tangency portfolio (5.26).

Example 5.4 *(Effect of β) Suppose the tangency portfolio has an expected excess return of 8% (which happens to be close to the value for the US market return since WWII). An asset with a beta of* 0.8 *should then have an expected excess return of* 6.4%*, and an asset*

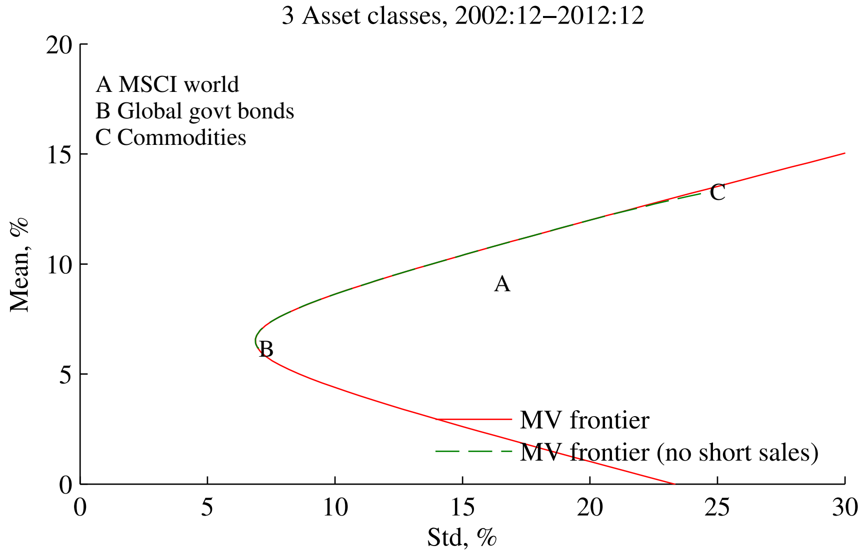


Figure 5.9: MV frontier, 3 asset classes

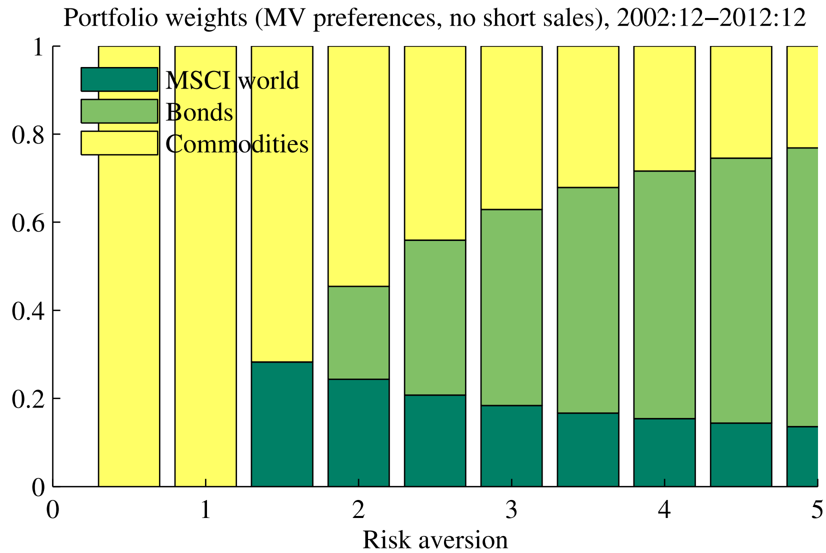


Figure 5.10: Portfolio choice (3 asset classes) with no short sales

with a beta of 1.2 *should have an expected excess return of* 9.6%.

Most stock indices (based on the standard characteristics like industry, size, value/growth) have betas around unity—but there are variations. For instance, building companies, manufacturers of investment goods and cars are typically often very procyclical (high betas), whereas food and drugs are not (low betas).

Proof. (of (5.31)) To derive 5.31, consider the asset 1 in the two asset case. We have

$$\text{Cov}(R_1, R_T) = \text{Cov}(R_1, w_1 R_1 + w_2 R_2) = w_1 \sigma_{11} + w_2 \sigma_{12}.$$

The expression for asset 2 is similar. Consider the first order conditions (5.8)–(5.9) for the investor with risk aversion k_T (for whom $v_i = w_i$)

$$\begin{aligned} \mu_1^e &= (w_1 \sigma_{11} + w_2 \sigma_{12}) k_T \\ &= \text{Cov}(R_1, w_1 R_1 + w_2 R_2) k_T \\ &= \text{Cov}(R_1, R_T) k_T. \end{aligned}$$

The expression for asset two is similar. Solve for the covariances as

$$\begin{aligned} \text{Cov}(R_1, R_T) &= \mu_1^e / k_T \\ \text{Cov}(R_2, R_T) &= \mu_2^e / k_T. \end{aligned}$$

These expressions will soon prove to be useful. Notice that the variance of the tangency portfolio is

$$\text{Var}(R_T) = \text{Cov}(w_1 R_1 + w_2 R_2, R_T) = w_1 \text{Cov}(R_1, R_T) + w_2 \text{Cov}(R_2, R_T),$$

which we can rewrite by using the expressions for the covariances above

$$\begin{aligned} \text{Var}(R_T) &= \left(w_1 \mu_1^e + w_2 \mu_2^e\right) / k_T \\ &= \mu_T^e / k_T. \end{aligned}$$

Consider asset 1. Divide $\text{Cov}(R_1, R_T)$ by $\text{Var}(R_T)$

$$\frac{\text{Cov}(R_1, R_T)}{\text{Var}(R_T)} = \frac{\mu_1^e / k_T}{\mu_T^e / k_T},$$

which can rearranged as (5.31). ■

Remark 5.5 *(Why is Risk = β? Short version) Because β measures the covariance with the market (and the idiosyncratic risk can be diversified away).*

Remark 5.6 *(Why is Risk = β? Longer Version) Start by investing 100% in the market portfolio, then increase position in asset i by a small amount (δ, 2% or so) by borrowing at the riskfree rate. The portfolio return is then*

$$R_p = R_m + \delta R_i^e.$$

The expected portfolio return is

$$\mathrm{E}\, R_p = \mathrm{E}\, R_m + \underbrace{\delta\, \mathrm{E}\, R_i^e}_{\textit{incremental risk premium}}$$

and the portfolio variance is

$$\mathrm{Var}(R_p) = \sigma_m^2 + \underbrace{\delta^2 \sigma_i^2 + 2\delta\, \mathrm{Cov}\,(R_i, R_m)}_{\textit{incremental risk, but } \delta^2\sigma_i^2 \approx 0}.$$

(For instance, if $\delta = 2\%$, then $\delta^2 = 0.0004$ and $2\delta = 0.04$.) Notice: risk = covariance with the market. The marginal compensation for more risk is

$$\frac{\textit{incremental risk premium}}{\textit{incremental risk}} = \frac{\mathrm{E}\, R_i^e}{2\,\mathrm{Cov}\,(R_i, R_m)}.$$

In equilibrium, the marginal compensation for more risk must be equal across assets

$$\frac{\mathrm{E}\, R_i^e}{2\,\mathrm{Cov}\,(R_i, R_m)} = \frac{\mathrm{E}\, R_j^e}{2\,\mathrm{Cov}\left(R_j, R_m\right)} = ... = \frac{\mathrm{E}\, R_m^e}{2\sigma_m^2},$$

since $\mathrm{Cov}\,(R_m, R_m) = \sigma_m^2$. *Rearrange as the CAPM expression.*

5.2.1 Beta of a Long-Short Position

Consider a zero cost portfolio consisting of one unit of asset i and minus one unit of asset j. The beta representation is clearly

$$\mu_i^e - \mu_j^e = \mathrm{E}(R_i - R_j) = (\beta_i - \beta_j)\mu_T^e. \tag{5.32}$$

If the two assets have the same betas, then this portfolio is not exposed to the tangency portfolio (and ought to carry a zero risk premium, at least according to theory). Such a long-short portfolio is a common way to isolate the investment from certain types of risk (here the systematic risk with respect to the tangency portfolio).

Proof. (of (5.32)) Notice that

$$\frac{\mathrm{Cov}\left(R_i - R_j, R_T\right)}{\mathrm{Var}\left(R_T\right)} = \frac{\mathrm{Cov}\left(R_i, R_T\right)}{\mathrm{Var}\left(R_T\right)} - \frac{\mathrm{Cov}\left(R_j, R_T\right)}{\mathrm{Var}\left(R_T\right)} = \beta_i - \beta_j.$$

■

5.3 Market Equilibrium

5.3.1 The Tangency Portfolio is the Market Portfolio

To determine the equilibrium asset prices (and therefore expected returns) we have to equate demand (the mean variance portfolios) with supply (exogenous). Since we assume a fixed and exogenous supply (say, 2000 shares of asset 1 and 407 shares of asset 2), prices (and therefore returns) are completely driven by demand.

Suppose all agents have the same beliefs about the asset returns (same expected returns and covariance mazrix). They will then all chose portfolios on the (same) efficient frontier—but possibly at different points (due to different risk aversions).

In equilibrium, net supply of the riskfree assets is zero (lending = borrowing), which implies that the optimal portfolio weights (5.12) must be such that the average (across investors) weights on the risky assets sum to unity ($v_1 + v_2 = 1$). These average values of v_1 and v_2, *the market portfolio*, then defines the tangency portfolio (denoted w_1 and w_2). In short, the tangency portfolio must be the market portfolio.

More formally, let the portfolio weights of investor j (with risk aversion k_j) be as in (5.22). Averaging across investors ($j = 1, 2, ..., J$) gives the average portfolio weights ($\bar{v}$, an $n \times 1$ vector)

$$\bar{v} = w \frac{1}{J} \sum\nolimits_{j=1}^{J} \frac{k_T}{k_j}. \tag{5.33}$$

This says that the average portfolio is proportional to the tangency portfolio (since all individual portfolios are). Summing across assets give the average position in the riskfree asset as

$$
\begin{aligned}
1 - \mathbf{1}'\bar{v} &= 1 - \mathbf{1}'w \frac{1}{J} \sum\nolimits_{j=1}^{J} \frac{k_T}{k_j} \\
&= 1 - \frac{1}{J} \sum\nolimits_{j=1}^{J} \frac{k_T}{k_j},
\end{aligned} \tag{5.34}
$$

since $\mathbf{1}'w$. This position should be zero, which identifies the risk aversion that is associated with the tangency portfolio as

$$
k_T = \frac{1}{\frac{1}{J} \sum_{j=1}^{J} \frac{1}{k_j}}. \tag{5.35}
$$

Clearly, when k_j is the same for all investors (so $k_T = k$), then they all hold the tangency portfolio.

Example 5.7 *("Average" risk aversion) If half of the investors have $k = 2$ and the other half has $k = 3$, then $k_T = 2.4$.*

(To simplify the notation, the previous analysis disregarded the possibility of different wealth levels of the investors. The extension is straightforward: instead of an unweighted average across investors, we need a weighted average where the weights reflect wealth relative to average wealth.)

5.3.2 Properties of the Market Portfolio

We can solve for μ_1^e and μ_2^e from the expressions for the optimal portfolio weights (5.12). In particular, do that for $k = k_T$ which we label k_m so $v = w$. In this case the portfolio weights are the same as in the market portfolio

$$
\begin{bmatrix} \mu_1^e \\ \mu_2^e \end{bmatrix} = k_m \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \tag{5.36}
$$

(or $\mu^e = k_m \Sigma w$ in matrix notation). Form the market (tangency) portfolio of the left hand side to get $\mathrm{E}\, R_m^e = w_1 \mu_1^e + w_2 \mu_2^e$. Forming the same portfolio of the right hand side gives $k_m \operatorname{Var}(R_m)$,

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}' \begin{bmatrix} \mu_1^e \\ \mu_2^e \end{bmatrix} = k_m \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}' \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \text{ or}$$

$$\mathrm{E}\, R_m^e = k_m \operatorname{Var}(R_m), \tag{5.37}$$

We can rearrange the last expression as

$$SR_m = \frac{\mathrm{E}\, R_m^e}{\operatorname{Std}(R_m)} = k_m \operatorname{Std}(R_m). \tag{5.38}$$

Since the tangency portfolio is the market portfolio, then this expression shows how the risk premium on the market is determined. The Sharpe ratio (5.38) is often called the "market price of risk." Having derived an expression for the risk premium, the asset prices can be calculated (not done here, since it is of little importance for our purposes).

Combining with the beta representation (5.31) we get

$$\begin{aligned} \mu_i^e &= \beta_i \,\mathrm{E}\, R_m^e \\ &= \beta_i k_m \operatorname{Var}(R_m). \end{aligned} \tag{5.39}$$

This shows that the expected excess return (risk premium) on asset i can be thought of as a product of three components: β_i which captures the covariance with the market, SR_m which is the price of market risk (risk compensation per unit of standard deviation of the market return), and $\operatorname{Std}(R_m)$ which measures the amount of market risk.

Notice that the expected return of asset i increases when *(i)* the riskfree rate increases; *(ii)* the market risk premium increases because of higher risk aversion or higher (beliefs about) market uncertainty; *(iii)* or when (beliefs about) beta increases.

An important feature of (5.39) is that the only movements in the return of asset i that matter for pricing are those movements that are correlated with the market (tangency portfolio) returns. In particular, if asset i and j have the same betas, then they have the same expected returns—even if one of them has a lot more uncertainty.

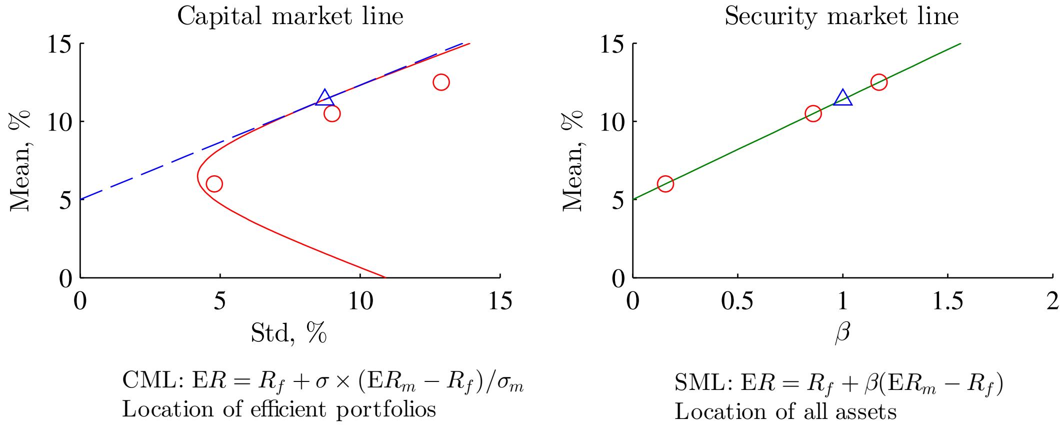


Figure 5.11: CML and SML

5.3.3 Summarizing MV and CAPM: CML and SML

According to MV analysis, all optimal portfolios (denoted opt) are on the *capital market line*

$$\mathrm{E}\, R_{opt} = R_f + \frac{\mathrm{E}\, R_m^e}{\mathrm{Std}(R_m)} \sigma_{opt}, \tag{5.40}$$

where $\mathrm{E}\, R_m^e$ and $\mathrm{Std}(R_m)$ are the expected value and the standard deviation of the excess return of the market portfolio. This is clearly the same as the upper leg of the MV frontier (with risky assets and riskfree asset). See Figure 5.11 for an example.

Proof. (of (5.40)) $R_{opt} = aR_m + (1-a)R_f$, so $R_{opt}^e = aR_m^e$. We then have $\mu_{opt}^e = a\mu_m^e$ and $\sigma_{opt} = a\sigma_m$ (since $a \geq 0$). Solve for a from the latter ($a = \sigma_{opt}/\sigma_m$) and use in the former. ∎

CAPM also implies that the beta representation (5.31) holds for any asset. Rewriting we have

$$\mu_i = R_f + \beta_i \,\mathrm{E}\, R_m^e. \tag{5.41}$$

The plot of μ_i against β_i (for different assets, i) is called the *security market line*. See Figure 5.11 for an example.

5.3.4 Back to Prices (Gordon Model)

The gross return, $1 + R_{t+1}$, is defined as

$$1 + R_{t+1} = \frac{D_{t+1} + P_{t+1}}{P_t}, \tag{5.42}$$

where P_t is the asset price and D_{t+1} the dividend it gives at the beginning of the next period.

Rearranging gives

$$P_t = \frac{D_{t+1}}{1 + R_{t+1}} + \frac{P_{t+1}}{1 + R_{t+1}}. \tag{5.43}$$

Use the same equation but with all time subscripts advanced one period ($P_{t+1} = \frac{D_{t+2}}{1+R_{t+2}} + \frac{P_{t+2}}{1+R_{t+2}}$) to substitute for P_{t+1}

$$P_t = \frac{D_{t+1}}{1 + R_{t+1}} + \frac{1}{1 + R_{t+1}} \left(\frac{D_{t+2}}{1 + R_{t+2}} + \frac{P_{t+2}}{1 + R_{t+2}} \right). \tag{5.44}$$

Now, substitute for P_{t+2} and then for P_{t+3} and so on. Finally, we have

$$P_t = \frac{D_{t+1}}{1 + R_{t+1}} + \frac{D_{t+2}}{(1 + R_{t+1})(1 + R_{t+2})} + \frac{D_{t+3}}{(1 + R_{t+1})(1 + R_{t+2})(1 + R_{t+3})} + \dots \tag{5.45}$$

$$= \sum_{j=1}^{\infty} \frac{D_{t+j}}{\prod_{s=1}^{j}(1 + R_{t+s})}. \tag{5.46}$$

We now make three simplifying assumptions. First, we can approximate the expectation of a ratio with the ratio of expectations ($\mathrm{E}(x/y) \approx \mathrm{E}\,x/\mathrm{E}\,y$). Second, that the expected j-period returns are $(1 + \mu)^j$

$$\mathrm{E}_t \prod_{s=1}^{j}(1 + R_{t+s}) \approx (1 + \mu)^j. \tag{5.47}$$

Third, that the expected dividends are constant $\mathrm{E}_t\, D_{t+j} = D$ and $\mathrm{E}_t\, R_{t+j} = \mu$ for all $j \geq 1$. We can then write (5.46) as

$$P_t \approx \sum_{j=1}^{\infty} \frac{D}{(1 + \mu)^j} = \frac{D}{\mu}, \tag{5.48}$$

which is clearly the Gordon model for an asset price.

If expected dividends increase, but expected returns do no (for instance, because the β of the asset is unchanged), then this is immediately capitalized in today's price (which increases). In contrast, if expected dividends are unchanged, but the expected (required) return increases, then today's asset price decreases.

5.4 An Application of MV Portfolio Choice: International Assets*

5.4.1 Foreign Investments

Let the exchange rate, S, be defined as units of domestic currency per unit of foreign currency, that is the price (measured in domestic currency) of foreign currency. Notice that a higher S means a weaker home currency (depreciation) and a lower S means a stronger home currency (appreciation).

Consider a *US investor buying British equity* in period *t*

$$\text{Investment}_{\$,t} = \text{Price of British equity}_{£,t} \times \text{ price of a GBP}_{\$,t} \tag{5.49}$$

...and selling in $t+1$

$$\text{Payoff}_{\$,t+1} = \text{Price of British equity}_{£,t+1} \times \text{ price of a GBP}_{\$,t+1} \tag{5.50}$$

The gross return, $1+R_u$, for US investor (in USD) is

$$\frac{\text{Payoff}_{\$,t+1}}{\text{Investment}_{\$,t}} = \underbrace{\frac{\text{Price of British equity}_{£,t+1}}{\text{Price of British equity}_{£,t}}}_{\text{local gross return}} \times \underbrace{\frac{\text{price of a GBP}_{\$,t+1}}{\text{price of a GBP}_{\$,t}}}_{\text{gross return on holding pounds}} \tag{5.51}$$

Simplify and approximate

$$\text{return in home currency} \approx \text{foreign (local) return + currency return} \tag{5.52}$$

Example 5.8 *(Investing abroad). The initial investment could have been*

5.5 *GBP per British share* $\times$ *1.6 USD per GBP = 8.8 USD,*

and the payoff

5.1 *GBP per British share* $\times$ *1.9 USD per GBP = 9.69 USD.*

The gross return can be written

$$1 + R_u = \frac{5.1}{5.5} \times \frac{1.9}{1.6} = (1 - 0.073) \times (1 + 0.188) = 1.10.$$

The approximation

$$R_u \approx -0.073 + 0.188 = 0.115$$

is not that bad.

To write the same in more general notation suppose we bought a foreign asset in t at the price P_t^*, measured in foreign currency; the cost in domestic currency was then $S_t P_t^*$. One period later (in $t + 1$), the value of the asset (in foreign currency) is P_{t+1}^* (think of this as the total value, including dividends or whatever); the value in domestic currency is thus $S_{t+1} P_{t+1}^*$. Clearly, the net return in domestic currency (unhedged), R_u, satisfies

$$\begin{aligned} 1 + R_u &= \frac{P_{t+1}^* S_{t+1}}{P_t^* S_t} && (5.53) \\ &= \frac{P_{t+1}^*}{P_t^*} \frac{S_{t+1}}{S_t} \\ &= (1 + R_*)(1 + R_s), && (5.54) \end{aligned}$$

where R_* is just the "local" return of the foreign asset (the return measured in foreign currency) and R_S is the return on the currency investment (buying foreign currency in t, selling it in $t + 1$) Notice that $R_s = S_{t+1}/S_t - 1$ is the percentage depreciation of the home currency (appreciation of the foreign currency). Someone who is investing abroad clearly benefits from the foreign currency becoming more expensive (the home currency becoming cheaper).

Clearly, we can rewrite the net return as

$$\begin{aligned} R_u &= R_* + R_s + R_s R_* && (5.55) \\ &\approx R_* + R_s && (5.56) \end{aligned}$$

where the approximation follows from the fact that the product of two net returns is typically very small (for instance, $0.05 \times 0.03 = 0.0015$). If we instead use log return (the log of the gross return), then there is no approximation error at all.

The approximation is used throughout this section (since it simplifies many expres-

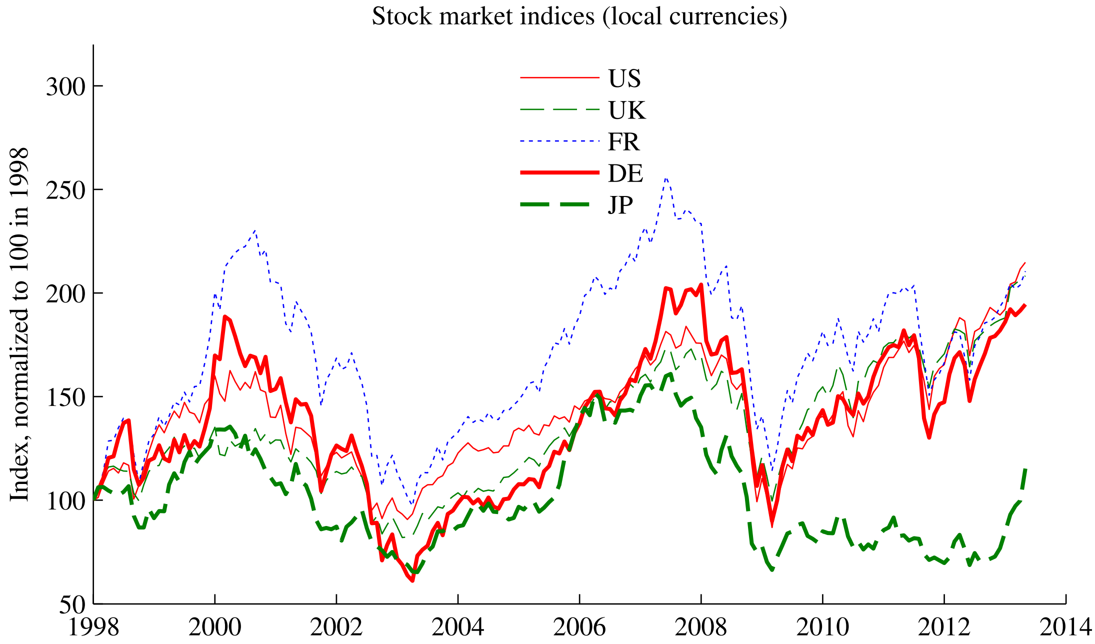


Figure 5.12: International stock market indices

sions considerably). The expected return and the variance (in domestic currency) are then

$$\mathrm{E}\, R_u \approx \mathrm{E}\, R_* + \mathrm{E}\, R_S, \text{ and} \tag{5.57}$$

$$\mathrm{Var}(R_u) \approx \mathrm{Var}(R_*) + \mathrm{Var}(R_s) + 2\,\mathrm{Cov}(R_s, R_*). \tag{5.58}$$

To apply the CAPM analysis to the problem of whether to invest internationally or not, suppose we have only two risky assets: a risky foreign equity index (with domestic currency return R_w) and a risky domestic equity index (denoted d). Then, according to (5.16) we should invest internationally if $\mu_w^e/\sigma_w > \rho\mu_d^e/\sigma_d$. This says that a high Sharpe ratio of the foreign asset (measured in domestic currency) or a low correlation with the domestic return both lead to investing internationally.

See Figures 5.12–5.13 and Tables 5.1–5.2 for an illustration.

Remark 5.9 *(Return from currency portfolios*) Buying foreign currency typically mean that you both buy that currency and then use that to pay for a foreign asset—often a foreign short-term debt instrument. Suppose you use 1 unit of domestic currency to buy $1/S_t^c$ units of foreign currency. You lend this foreign currency at the interest rate i^c, so one*

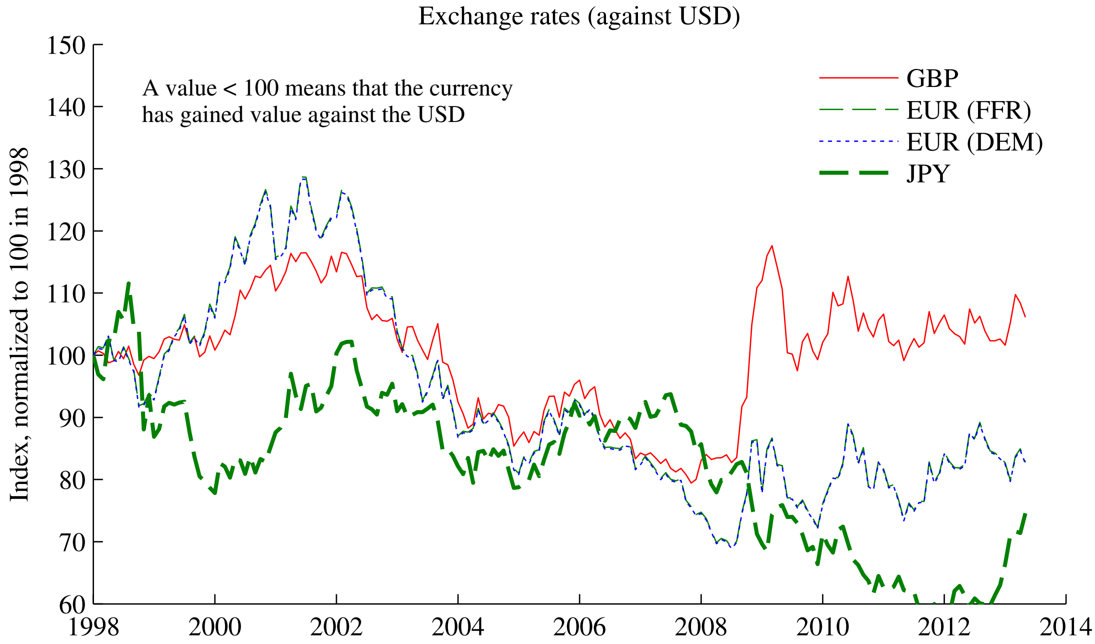


Figure 5.13: Exchange rate indices

	Local currency	Exchange rate	in USD
US	6.6	0.0	6.6
UK	6.0	0.0	6.0
FR	6.8	1.8	8.6
DE	7.0	1.8	8.9
JP	2.8	2.5	5.4

Table 5.1: Contribution to the average return for a US investor investing in different equity markets, 1998:1-2013:5

period later you have $(1+i^c)/S^c_t$ *units of foreign currency, which you sell at the exchange rate* S^c_{t+1} *to get domestic currency. Your net return is* $S^c_{t+1}(1+i^c)/S^c_t - 1$. *If you financed his investment by borrowing on the domestic money market at the interest rate* i, *then the excess return of your investment in country* c *was* $R^e_c = [S^c_{t+1}(1+i^c)/S^c_t - 1] - i$. *In many cases, this is approximated as* $\ln(S^c_{t+1}/S^c_t) + (i^c - i)$, *where the first term is the depreciation of the domestic currency (that is, the appreciation of the foreign currency) and the second term is the interest rate differential.*

	Local currency	Exchange rate	2*Cov	in USD
US	3.1	0.0	0.0	3.1
UK	2.3	0.8	0.3	3.4
FR	3.8	1.2	0.5	5.4
DE	5.3	1.2	0.5	7.0
JP	3.8	1.3	−1.5	3.6

Table 5.2: Contribution to the variance of the return for a US investor investing in different equity markets, 1998:1-2013:5

Remark 5.10 *(Return for carry trade portfolios*) Now, for another country* (d) *you might reverse these positions, and the excess return becomes* $R_d^e = -[S_{t+1}^d(1+i^d)/S_t^d - 1] + i$ *which is approximately* $\ln(S_{t+1}^d/S_t^d) + (i^d - i)$. *Clearly, you can put these positions together in carry trade one portfolio to have* $R_c^e + R_d^e = S_{t+1}^c(1+i^c)/S_t^c - S_{t+1}^d(1+i^d)/S_t^d$, *which is approximately* $\ln[(S_{t+1}^c/S_{t+1}^d)/(S_t^c/S_t^d)] + (i^c - i^d)$. *Since* S_t^c/S_t^d *is the cross rate (number of currency* c *units that you pay to buy one unit of currency* d*), the approximate expression includes appreciation of currency* c *relative to currency* d *plus their interest rate differential. (This is very close to explicitly borrowing currency* d *to buy* c *and lend there.)*

5.4.2 Invest in Foreign Stocks? Rule-of-Thumb

The result in (5.17) provides a simple rule of thumb for whether we should invest in foreign assets or not. Let asset 1 represent a domestic market index, and asset 2 a foreign market index. The rule is then: invest in the foreign market if its Sharpe ratio is higher than the Sharpe ratio of the domestic market times the correlation of the two markets (that is, if $\mu_2^e/\sigma_2 > \rho\mu_1^e/\sigma_1$). Clearly, the returns should be measured in the same currency (but the currency risk may be hedged or not).

See Figure 5.14 for an example.

5.5 Testing CAPM

Reference: Elton, Gruber, Brown, and Goetzmann (2010) 15

Let $R_{it}^e = R_{it} - R_{ft}$ be the excess return on asset i in excess over the riskfree asset, and let R_{mt}^e be the excess return on the market portfolio. The basic implication of CAPM

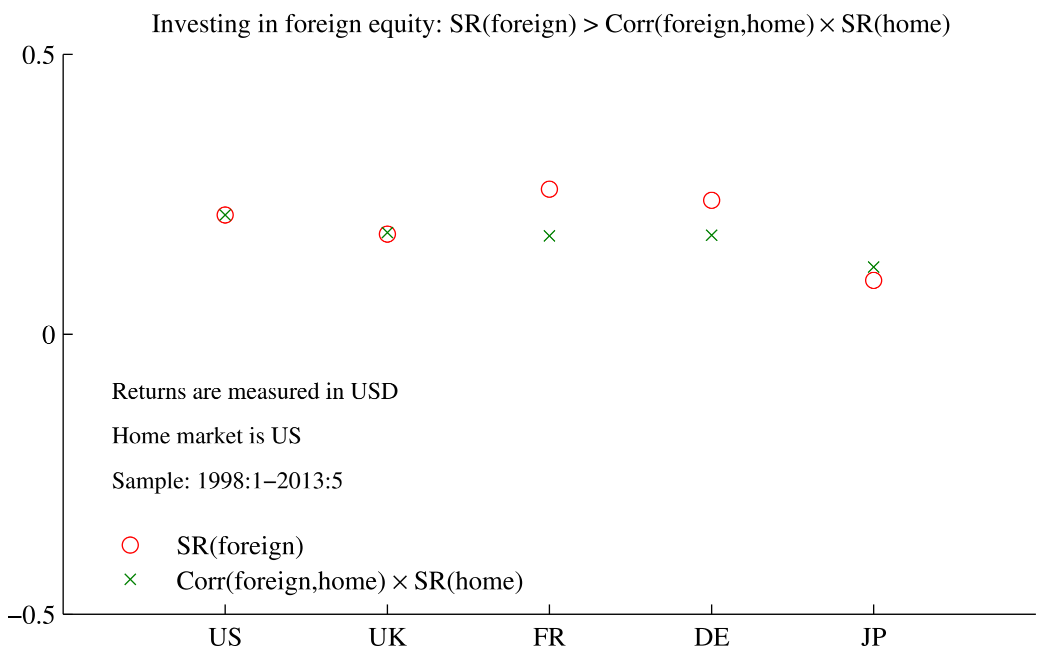


Figure 5.14: International stock indices

is that the expected excess return of an asset ($\mathrm{E}\, R^e_{it}$) is linearly related to the expected excess return on the market portfolio ($\mathrm{E}\, R^e_{mt}$) according to

$$\mathrm{E}\, R^e_{it} = \beta_i \,\mathrm{E}\, R^e_{mt}, \text{ where } \beta_i = \frac{\mathrm{Cov}\,(R_i, R_m)}{\mathrm{Var}\,(R_m)}. \tag{5.59}$$

Consider the regression

$$\begin{aligned} R^e_{it} &= \alpha_i + b_i R^e_{mt} + \varepsilon_{it}, \text{ where} \\ \mathrm{E}\, \varepsilon_{it} &= 0 \text{ and } \mathrm{Cov}(R^e_{mt}, \varepsilon_{it}) = 0. \end{aligned} \tag{5.60}$$

The two last conditions are automatically imposed by LS. Take expectations of the regression to get

$$\mathrm{E}\, R^e_{it} = \alpha_i + b_i \,\mathrm{E}\, R^e_{mt}. \tag{5.61}$$

Notice that the LS estimate of b_i is the sample analogue to β_i in (5.59). It is then clear that CAPM implies that the intercept (α_i) of the regression should be zero, which is also what empirical tests of CAPM focus on.

This test of CAPM can be given two interpretations. If we assume that R_{mt} is the correct benchmark (the tangency portfolio for which (5.59) is true by definition), then it is a test of whether asset R_{it} is correctly priced. This is typically the perspective in performance analysis of mutual funds. Alternatively, if we assume that R_{it} is correctly priced, then it is a test of the mean-variance efficiency of R_{mt}. This is the perspective of CAPM tests.

The t-test of the null hypothesis that $\alpha_i = 0$ uses the fact that, under fairly mild conditions, the t-statistic has an asymptotically normal distribution, that is

$$\frac{\hat{\alpha}_i}{\text{Std}(\hat{\alpha}_i)} \xrightarrow{d} N(0,1) \text{ under } H_0 : \alpha_i = 0. \tag{5.62}$$

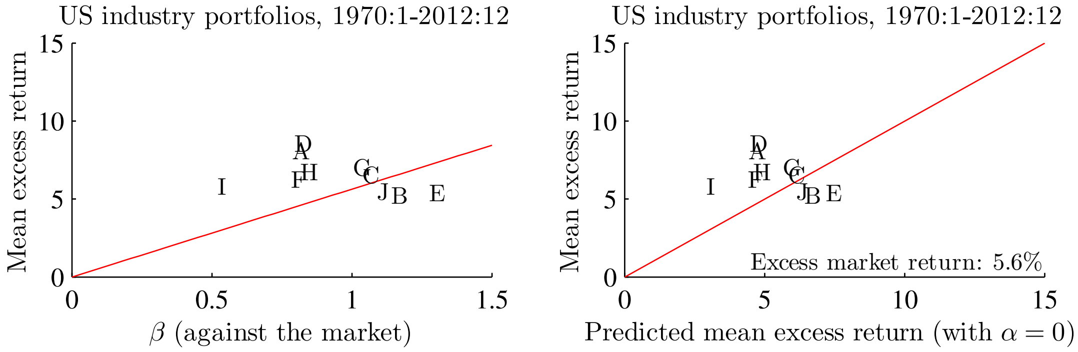
Note that this is the distribution under the null hypothesis that the true value of the intercept is zero, that is, that CAPM is correct (in this respect, at least).

The test assets are typically portfolios of firms with similar characteristics, for instance, small size or having their main operations in the retail industry. There are two main reasons for testing the model on such portfolios: individual stocks are extremely volatile and firms can change substantially over time (so the beta changes). Moreover, it is of interest to see how the deviations from CAPM are related to firm characteristics (size, industry, etc), since that can possibly suggest how the model needs to be changed.

The results from such tests vary with the test assets used. For US portfolios, CAPM seems to work reasonably well for some types of portfolios (for instance, portfolios based on firm size or industry), but much worse for other types of portfolios (for instance, portfolios based on firm dividend yield or book value/market value ratio). Figure 5.15 shows some results for US industry portfolios.

5.5.1 Econometric Properties of the CAPM Test

A common finding from Monte Carlo simulations is that these tests tend to reject a true null hypothesis too often when the critical values from the asymptotic distribution are used: the actual small sample *size of the test* is thus larger than the asymptotic (or "nominal") size (see Campbell, Lo, and MacKinlay (1997) Table 5.1). The practical consequence is that we should either used adjusted critical values (from Monte Carlo or bootstrap simulations)—or more pragmatically, that we should only believe in strong rejections of the null hypothesis.



	alpha	pval	StdErr
all	NaN	0.04	NaN
A (NoDur)	3.62	0.01	8.70
B (Durbl)	-1.21	0.55	13.66
C (Manuf)	0.70	0.48	6.37
D (Enrgy)	4.06	0.07	14.75
E (HiTec)	-1.82	0.32	11.93
F (Telcm)	1.82	0.29	11.10
G (Shops)	1.37	0.35	9.51
H (Hlth)	2.13	0.21	11.39
I (Utils)	2.87	0.11	11.64
J (Other)	-0.65	0.55	6.99

CAPM
Factor: US market
alpha and StdErr are in annualized %

Figure 5.15: CAPM regressions on US industry indices

To study the power of the test (the frequency of rejections of a false null hypothesis) we have to specify an alternative data generating process (for instance, how much extra return in excess of that motivated by CAPM) and the size of the test (the critical value to use). Once that is done, it is typically found that these tests require a substantial deviation from CAPM and/or a long sample to get good power. The basic reason for this is that asset returns are very volatile. For instance, suppose that the standard OLS assumptions (iid residuals that are independent of the market return) are correct. Then, it is straightforward to show that the variance of Jensen's alpha is

$$\text{Var}(\hat{\alpha}_i) = \left[1 + \frac{(\mu_m^e)^2}{\text{Var}\left(R_m^e\right)}\right]\sigma^2/T \tag{5.63}$$

$$= [1 + (SR_m)^2]\sigma^2/T, \tag{5.64}$$

where σ^2 is the variance of the residual in (5.60) and SR_m is the Sharpe ratio of the market portfolio. We see that the uncertainty about the alpha is high when the residual is volatile and when the sample is short, but also when the Sharpe ratio of the market is high. Note that a large market Sharpe ratio means that the market asks for a high compensation for taking on risk. A bit uncertainty about how risky asset i is then translates in a large uncertainty about what the risk-adjusted return should be.

Example 5.11 *Suppose we have monthly data with* $\widehat{\alpha}_i = 0.2\%$ *(that is,* $0.2\% \times 12 = 2.4\%$ *per year),* $\sigma = 3\%$ *(that is,* $3\% \times \sqrt{12} \approx 10\%$ *per year) and a market Sharpe ratio of* 0.15 *(that is,* $0.15 \times \sqrt{12} \approx 0.5$ *per year). (This corresponds well to US CAPM regressions for industry portfolios.) A significance level of* 10% *requires a t-statistic (5.62) of at least 1.65, so*

$$\frac{0.2}{\sqrt{1+0.15^2}3/\sqrt{T}} \geq 1.65 \text{ or } T \geq 626.$$

We need a sample of at least 626 months (52 years)! With a sample of only 26 years (312 months), the alpha needs to be almost 0.3% per month (3.6% per year) or the standard deviation of the residual just 2% *(7% per year). Notice that cumulating a 0.3% return over 25 years means almost 2.5 times the initial value.*

Proof. (*Proof of (5.64)) Consider the regression equation $y_t = x_t'b + \varepsilon_t$. With iid errors that are independent of all regressors (also across observations), the LS estimator, $\hat{b}_{Ls}$, is asymptotically distributed as

$$\sqrt{T}(\hat{b}_{Ls} - b) \overset{d}{\to} N(\mathbf{0}, \sigma^2 \Sigma_{xx}^{-1}), \text{ where } \sigma^2 = \mathrm{Var}(\varepsilon_t) \text{ and } \Sigma_{xx} = \mathrm{plim}\Sigma_{t=1}^T x_t x_t'/T.$$

When the regressors are just a constant (equal to one) and one variable regressor, f_t, so $x_t = [1, f_t]'$, then we have

$$\Sigma_{xx} = \mathrm{E} \textstyle\sum_{t=1}^T x_t x_t'/T = \mathrm{E}\frac{1}{T}\sum_{t=1}^T \begin{bmatrix} 1 & f_t \\ f_t & f_t^2 \end{bmatrix} = \begin{bmatrix} 1 & \mathrm{E}\, f_t \\ \mathrm{E}\, f_t & \mathrm{E}\, f_t^2 \end{bmatrix}, \text{ so}$$

$$\sigma^2 \Sigma_{xx}^{-1} = \frac{\sigma^2}{\mathrm{E}\, f_t^2 - (\mathrm{E}\, f_t)^2} \begin{bmatrix} \mathrm{E}\, f_t^2 & -\mathrm{E}\, f_t \\ -\mathrm{E}\, f_t & 1 \end{bmatrix} = \frac{\sigma^2}{\mathrm{Var}(f_t)} \begin{bmatrix} \mathrm{Var}(f_t) + (\mathrm{E}\, f_t)^2 & -\mathrm{E}\, f_t \\ -\mathrm{E}\, f_t & 1 \end{bmatrix}.$$

(In the last line we use $\mathrm{Var}(f_t) = \mathrm{E}\, f_t^2 - (\mathrm{E}\, f_t)^2$.) ■

5.5.2 Several Assets

In most cases there are several (n) test assets, and we actually want to test if all the α_i (for $i = 1, 2, ..., n$) are zero. Ideally we then want to take into account the correlation of the different alphas.

While it is straightforward to construct such a test, it is also a bit messy. As a quick way out, the following will work fairly well. First, test each asset individually. Second, form a few different portfolios of the test assets (equally weighted, value weighted) and test these portfolios. Although this does not deliver one single test statistic, it provides plenty of information to base a judgement on. For a more formal approach, a SURE approach is useful.

A quite different approach to study a cross-section of assets is to first perform a CAPM regression (5.60) and then the following cross-sectional regression

$$\sum_{t=1}^{T} R_{it}^e / T = \gamma + \lambda \hat{\beta}_i + u_i, \tag{5.65}$$

where $\sum_{t=1}^{T} R_{it}^e / T$ is the (sample) average excess return on asset i. Notice that the estimated betas are used as regressors and that there are as many data points as there are assets (n).

There are severe econometric problems with this regression equation since the regressor contains measurement errors (it is only an uncertain estimate), which typically tend to bias the slope coefficient towards zero. To get the intuition for this bias, consider an extremely noisy measurement of the regressor: it would be virtually uncorrelated with the dependent variable (noise isn't correlated with anything), so the estimated slope coefficient would be close to zero.

If we could overcome this bias (and we can by being careful), then the testable implications of CAPM is that $\gamma = 0$ and that λ equals the average market excess return. We also want (5.65) to have a high R^2—since it should be unity in a very large sample (if CAPM holds).

5.5.3 Representative Results of the CAPM Test

One of the more interesting studies is Fama and French (1993) (see also Fama and French (1996)). They construct 25 stock portfolios according to two characteristics of the firm:

the size (by market capitalization) and the book-value-to-market-value ratio (BE/ME). In June each year, they sort the stocks according to size and BE/ME. They then form a 5×5 matrix of portfolios, where portfolio ij belongs to the ith size quintile *and* the jth BE/ME quintile:

$$\begin{bmatrix} \text{small size, low B/M} & \ldots & \ldots & \ldots & \text{small size, high B/M} \\ \vdots & \ddots & & & \\ \vdots & & \ddots & & \\ \vdots & & & \ddots & \\ \text{large size, low B/M} & & & & \text{large size, high B/M} \end{bmatrix}$$

Tables 5.3–5.4 summarize some basic properties of these portfolios.

	Book value/Market value				
	1	2	3	4	5
Size 1	3.3	9.2	9.6	11.7	13.2
2	5.4	8.4	10.5	10.8	12.0
3	5.7	8.9	8.8	10.3	12.0
4	6.8	6.7	8.6	9.7	9.6
5	5.2	5.8	6.1	5.9	7.3

Table 5.3: Mean excess returns (annualised %), US data 1957:1–2012:12. Size 1: smallest 20% of the stocks, Size 5: largest 20% of the stocks. B/M 1: the 20% of the stocks with the smallest ratio of book to market value (growth stocks). B/M 5: the 20% of the stocks with the highest ratio of book to market value (value stocks).

They run a traditional CAPM regression on each of the 25 portfolios (monthly data 1963–1991)—and then study if the expected excess returns are related to the betas as they should according to CAPM (recall that CAPM implies $\mathrm{E}\, R^e_{it} = \beta_i \lambda$ where λ is the risk premium (excess return) on the market portfolio).

However, it is found that there is almost no relation between $\mathrm{E}\, R^e_{it}$ and β_i (there is a cloud in the $\beta_i \times \mathrm{E}\, R^e_{it}$ space, see Cochrane (2001) 20.2, Figure 20.9). This is due to the combination of two features of the data. First, *within a BE/ME quintile*, there is a positive relation (across size quantiles) between $\mathrm{E}\, R^e_{it}$ and β_i—as predicted by CAPM (see Cochrane (2001) 20.2, Figure 20.10). Second, *within a size quintile* there is a negative relation (across BE/ME quantiles) between $\mathrm{E}\, R^e_{it}$ and β_i—in stark contrast to CAPM (see Cochrane (2001) 20.2, Figure 20.11).

	Book value/Market value				
	1	2	3	4	5
Size 1	1.4	1.2	1.1	1.0	1.1
2	1.4	1.2	1.0	1.0	1.1
3	1.3	1.1	1.0	1.0	1.0
4	1.2	1.1	1.0	1.0	1.0
5	1.0	0.9	0.9	0.8	0.9

Table 5.4: Beta against the market portfolio, US data 1957:1–2012:12. Size 1: smallest 20% of the stocks, Size 5: largest 20% of the stocks. B/M 1: the 20% of the stocks with the smallest ratio of book to market value (growth stocks). B/M 5: the 20% of the stocks with the highest ratio of book to market value (value stocks).

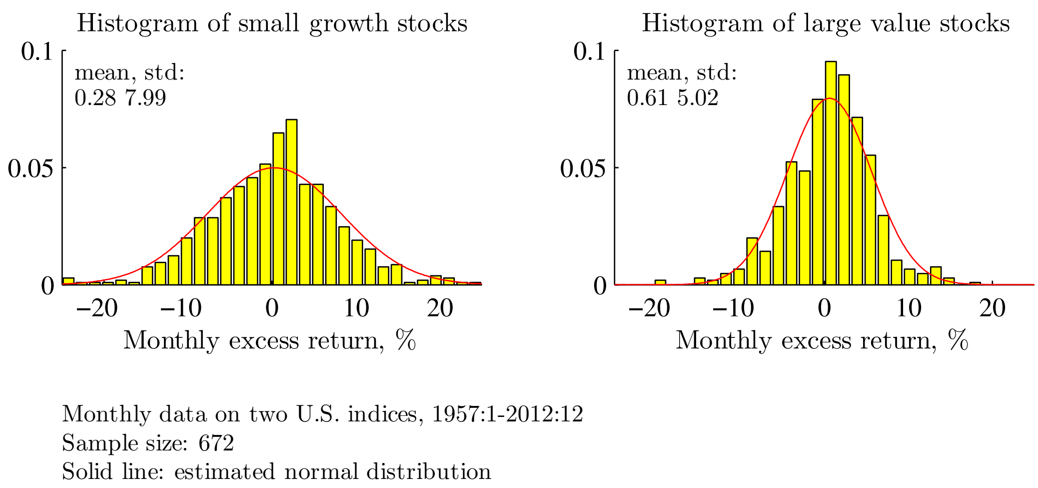


Figure 5.16: Comparison of small growth stock and large value stocks

Figure 5.15 shows some results for US industry portfolios and Figures 5.17–5.19 for US size/book-to-market portfolios.

5.5.4 Representative Results on Mutual Fund Performance

Mutual fund evaluations (estimated α_i) typically find *(i)* on average neutral performance (or less: trading costs&fees); *(ii)* large funds might be worse; *(iii)* perhaps better performance on less liquid (less efficient?) markets; and *(iv)* there is very little persistence in performance: α_i for one sample does not predict α_i for subsequent samples (except for

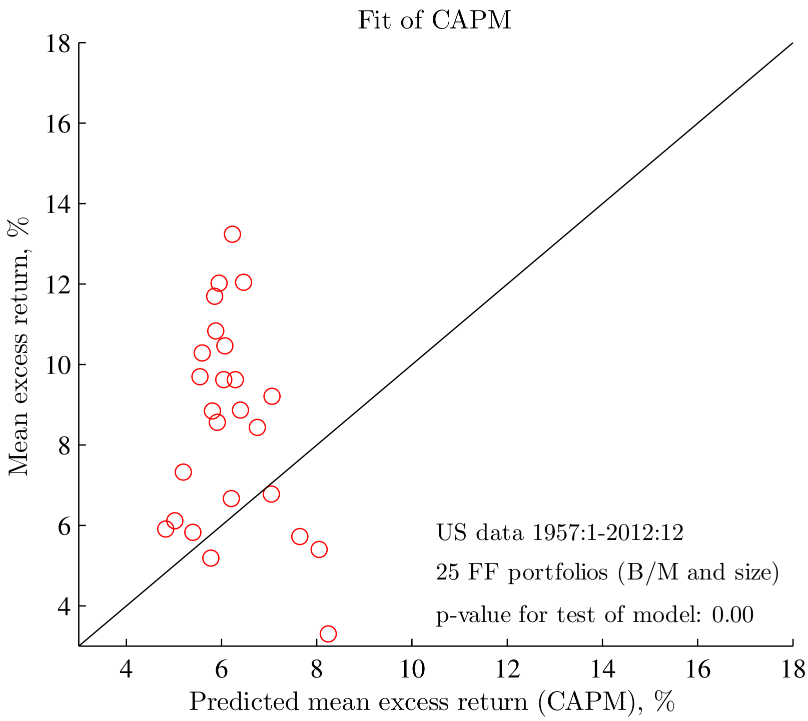


Figure 5.17: CAPM, FF portfolios

bad funds).

A Statistical Tables

Bibliography

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, 1997, *The econometrics of financial markets*, Princeton University Press, Princeton, New Jersey.

Cochrane, J. H., 2001, *Asset pricing*, Princeton University Press, Princeton, New Jersey.

Danthine, J.-P., and J. B. Donaldson, 2002, *Intermediate financial theory*, Prentice Hall.

Elton, E. J., M. J. Gruber, S. J. Brown, and W. N. Goetzmann, 2010, *Modern portfolio theory and investment analysis*, John Wiley and Sons, 8th edn.

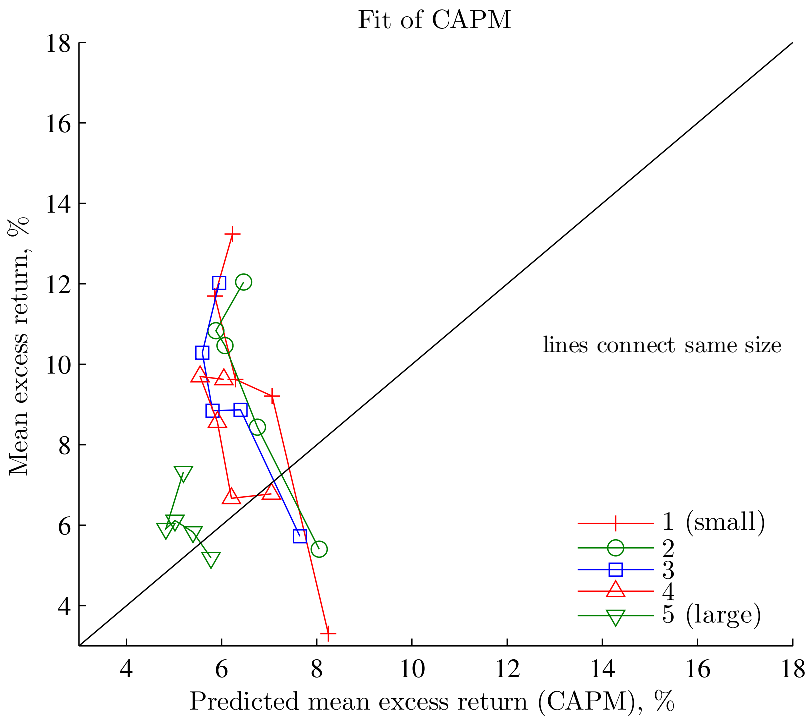


Figure 5.18: CAPM, FF portfolios

n	Critical values		
	10%	5%	1%
10	1.81	2.23	3.17
20	1.72	2.09	2.85
30	1.70	2.04	2.75
40	1.68	2.02	2.70
50	1.68	2.01	2.68
60	1.67	2.00	2.66
70	1.67	1.99	2.65
80	1.66	1.99	2.64
90	1.66	1.99	2.63
100	1.66	1.98	2.63
Normal	1.64	1.96	2.58

Table A.1: Critical values (two-sided test) of t distribution (different degrees of freedom) and normal distribution.

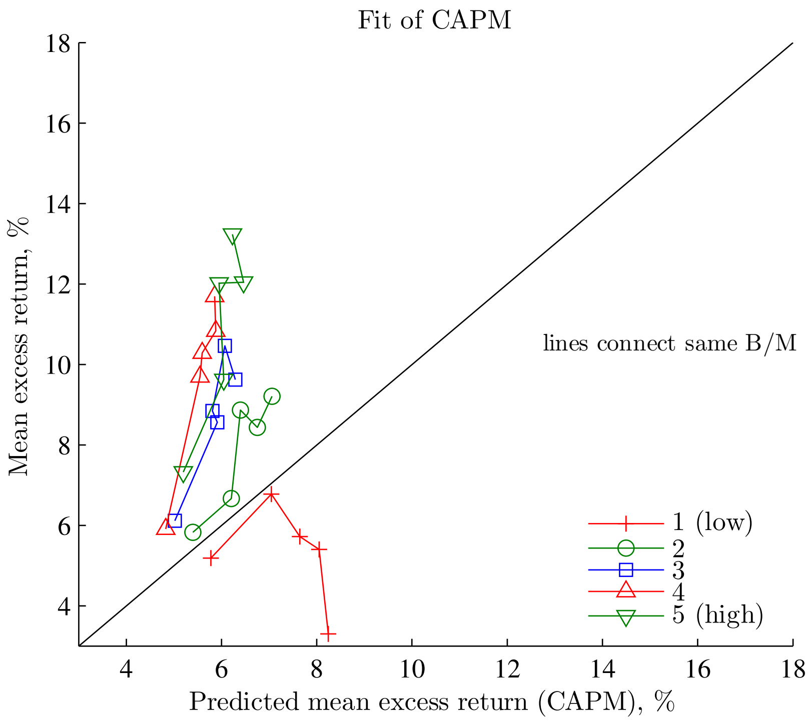


Figure 5.19: CAPM, FF portfolios

n	Critical values		
	10%	5%	1%
1	2.71	3.84	6.63
2	4.61	5.99	9.21
3	6.25	7.81	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48
8	13.36	15.51	20.09
9	14.68	16.92	21.67
10	15.99	18.31	23.21

Table A.2: Critical values of chisquare distribution (different degrees of freedom, n).

Fama, E. F., and K. R. French, 1993, “Common risk factors in the returns on stocks and bonds,” *Journal of Financial Economics*, 33, 3–56.

Fama, E. F., and K. R. French, 1996, "Multifactor explanations of asset pricing anomalies," *Journal of Finance*, 51, 55–84.

6 Performance Analysis

Reference: Elton, Gruber, Brown, and Goetzmann (2010) 25
More advanced material is denoted by a star (*). It is not required reading.

6.1 Performance Evaluation

Reference: Elton, Gruber, Brown, and Goetzmann (2010) 25

6.1.1 The Idea behind Performance Evaluation

Traditional performance analysis tries to answer the following question: "should we include an asset in our portfolio, assuming that future returns will have the same distribution as in a historical sample." Since returns are random variables (although with different means, variances, etc) and investors are risk averse, this means that performance analysis will typically *not* rank the fund with the highest return (in a historical sample) first. Although that high return certainly was good for the old investors, it is more interesting to understand what kind of distribution of future returns this investment strategy might entail. In short, the high return will be compared with the risk of the strategy.

Most performance measures are based on mean-variance analysis, but the full MV portfolio choice problem is not solved. Instead, the performance measures can be seen as different approximations of the MV problem, where the issue is whether we should invest in fund p or in fund q. (We don't allow a mix of them.) Although the analysis is based on the MV model, it is not assumed that all assets (portfolios) obey CAPM's beta representation—or that the market portfolio must be the optimal portfolio for every investor. One motivation of this approach could be that the investor (who is doing the performance evaluation) is a MV investor, but that the market is influenced by non-MV investors.

Of course, the analysis is also based on the assumption that historical data are good forecasters of the future.

There are several popular performance measures, corresponding to different situations: is this an investment of your entire wealth, or just a small increment? However, all these measures are (increasing) functions of Jensen's alpha, the intercept in the CAPM regression

$$\begin{aligned} R_{it}^e &= \alpha_i + b_i R_{mt}^e + \varepsilon_{it}, \text{ where} \\ \mathrm{E}\,\varepsilon_{it} &= 0 \text{ and } \mathrm{Cov}(R_{mt}^e, \varepsilon_{it}) = 0. \end{aligned} \tag{6.1}$$

Example 6.1 *(Statistics for example of performance evaluations) We have the following information about portfolios m (the market), p, and q*

	α	β	Std(ε)	μ^e	σ
m	0.000	1.000	0.000	0.100	0.180
p	0.010	0.900	0.140	0.100	0.214
q	0.050	1.300	0.030	0.180	0.236

Table 6.1: Basic facts about the market and two other portfolios, α, β, and Std(ε) are from CAPM regression: $R_{it}^e = \alpha + \beta R_{mt}^e + \varepsilon_{it}$

6.1.2 Sharpe Ratio and M^2: Evaluating the Overall Portfolio

Suppose we want to know if fund p is better than fund q to place *all* our savings in. (We don't allow a mix of them.) The answer is that p is better if it has a higher Sharpe ratio—defined as

$$SR_p = \mu_p^e / \sigma_p. \tag{6.2}$$

The reason is that MV behaviour (MV preferences or normally distributed returns) implies that we should maximize the Sharpe ratio (selecting the tangency portfolio). Intuitively, for a given volatility, we then get the highest expected return.

Example 6.2 *(Performance measure) From Example 6.1 we get the following performance measures*

A version of the Sharpe ratio, called M^2 (after some of the early proponents of the measure: Modigliani and Modigliani) is

$$M_p^2 = \mu_{p*}^e - \mu_m^e \text{ (or } \mu_{p*} - \mu_m), \tag{6.3}$$

	SR	M^2	AR	Treynor	T^2
m	0.556	0.000		0.100	0.000
p	0.467	−0.016	0.071	0.111	0.011
q	0.763	0.037	1.667	0.138	0.038

Table 6.2: Performance Measures

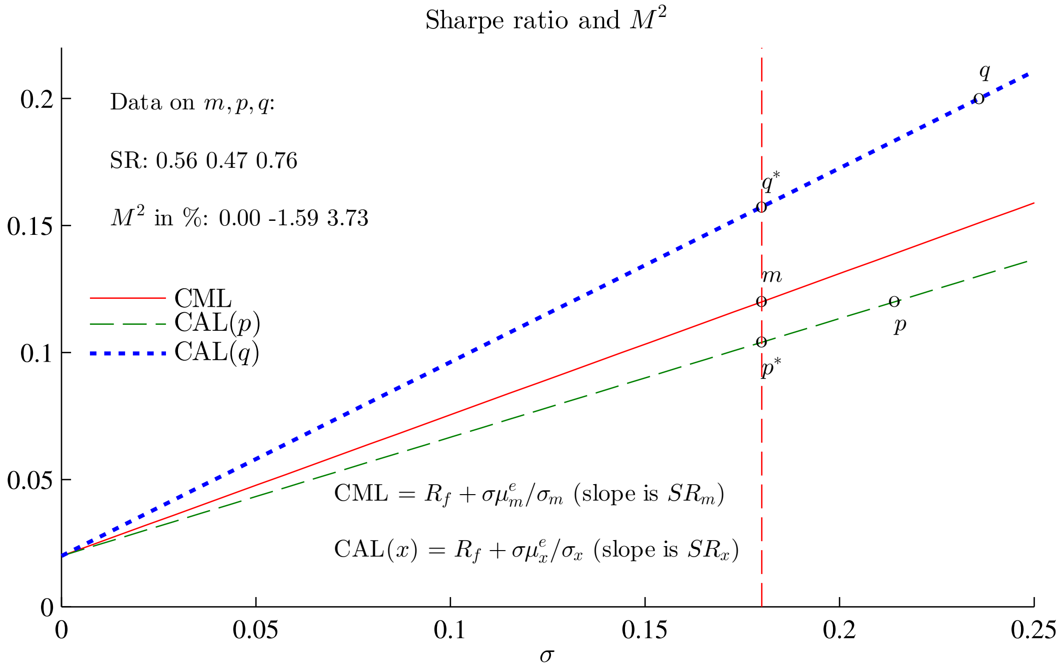


Figure 6.1: Sharpe ratio and M^2

where $\mu_{p^*}^e$ is the expected return on a mix of portfolio p and the riskfree asset such that the volatility is the same as for the market return.

$$R_{p^*} = aR_p + (1-a)R_f, \text{ with } a = \sigma_m/\sigma_p. \tag{6.4}$$

This gives the mean and standard deviation of portfolio p^*

$$\mu_{p^*}^e = a\mu_p^e = \mu_p^e\sigma_m/\sigma_p \tag{6.5}$$

$$\sigma_{p^*} = a\sigma_p = \sigma_m. \tag{6.6}$$

The latter shows that R_{p^*} indeed has the same volatility as the market. See Example 6.2 and Figure 6.1 for an illustration.

M^2 has the advantage of being easily interpreted—it is just a comparison of two returns. It shows how much better (or worse) this asset is compared to the capital market line (which is the location of efficient portfolios provided the market is MV efficient). However, it is just a scaling of the Sharpe ratio.

To see that, use (6.2) to write

$$\begin{aligned} M_p^2 &= SR_{p^*}\sigma_{p^*} - SR_m\sigma_m \\ &= \left(SR_p - SR_m\right)\sigma_m. \end{aligned} \tag{6.7}$$

The second line uses the facts that R_{p^*} has the same Sharpe ratio as R_p (see (6.5)–(6.6)) and that R_{p^*} has the same volatility as the market. Clearly, the portfolio with the highest Sharpe ratio has the highest M^2.

6.1.3 Appraisal Ratio: Which Portfolio to Combine with the Market Portfolio?

If the issue is "should I *add* fund p or fund q to my holding of the market portfolio?," then the appraisal ratio provides an answer. The appraisal ratio of fund p is

$$AR_p = \alpha_p / \operatorname{Std}(\varepsilon_{pt}), \tag{6.8}$$

where α_p is the intercept and $\operatorname{Std}(\varepsilon_{pt})$ the volatility of the residual of a CAPM regression (6.1). (The residual is often called the tracking error.) A higher appraisal ratio is better.

If you think of $b_p R^e_{mt}$ as the benchmark return, then AR_p is the average extra ruturn per unit of extra volatility (standard deviation). For instance, a ration of 1.7 could be interpreted as a 1.7 USD profit per each dollar risked.

The motivation is that if we take the market portfolio and portfolio p to be the available assets, and then find the optimal (assuming MV preferences) combination of them, then the squared Sharpe ratio of the optimal portfolio (that is, the tangency portfolio) is

$$SR_c^2 = \left(\frac{\alpha_p}{\operatorname{Std}(\varepsilon_{pt})}\right)^2 + SR_m^2. \tag{6.9}$$

If the alpha is positive, a higher appraisal ratio gives a higher Sharpe ratio—which is the objective if we have MV preferences. See Example 6.2 for an illustration.

If the alpha is negative, and we rule out short sales, then (6.9) is less relevant. In this case, the optimal portfolio weight on an asset with a negative alpha is (very likely to be) zero—so those assets are uninteresting.

The *information ratio*

$$IR_p = \frac{\mathrm{E}(R_p - R_b)}{\mathrm{Std}(R_p - R_b)}, \tag{6.10}$$

where R_b is some benechmark return is similar to the appraisal ratio—although a bit more general. In the information ratio, the denominator can be thought of as the tracking error relative to the benchmark—and the numerator as the gain from deviating. Notice, however, that when the benchmark is $b_p R^e_{mt}$, then the information ratio is the same as the appraisal ratio.

Proof. From the CAPM regression (6.1) we have

$$\mathrm{Cov}\begin{bmatrix} R^e_{it} \\ R^e_{mt} \end{bmatrix} = \begin{bmatrix} \beta_i^2\sigma_m^2 + \mathrm{Var}(\varepsilon_{it}) & \beta_i\sigma_m^2 \\ \beta_i\sigma_m^2 & \sigma_m^2 \end{bmatrix}, \text{ and } \begin{bmatrix} \mu^e_i \\ \mu^e_m \end{bmatrix} = \begin{bmatrix} \alpha_i + \beta_i\mu^e_m \\ \mu^e_m \end{bmatrix}.$$

Suppose we use this information to construct a mean-variance frontier for both R_{it} and R_{mt}, and we find the tangency portfolio, with excess return R^e_{ct}. We assume that there are no restrictions on the portfolio weights. Recall that the square of the Sharpe ratio of the tangency portfolio is $\mu^{e\prime}\Sigma^{-1}\mu^e$, where μ^e is the vector of expected excess returns and Σ is the covariance matrix. By using the covariance matrix and mean vector above, we get that the squared Sharpe ratio for the tangency portfolio (using both R_{it} and R_{mt}) is

$$\left(\frac{\mu^e_c}{\sigma_c}\right)^2 = \frac{\alpha_i^2}{\mathrm{Var}(\varepsilon_{it})} + \left(\frac{\mu^e_m}{\sigma_m}\right)^2.$$

■

6.1.4 Treynor's Ratio and T^2: Portfolio is a Small Part of the Overall Portfolio

Suppose instead that the issue is if we should add a *small* amount of fund p or fund q to an already well diversified portfolio (not the market portfolio). In this case, Treynor's ratio might be useful

$$TR_p = \mu^e_p/\beta_p. \tag{6.11}$$

A higher Treynor's ratio is better.

The TR measure can be rephrased in terms of expected returns—and could then called the T^2 measure. Mix p and q with the riskfree rate to get the same β for both portfolios (here 1 to make it comparable with the market), the one with the highest Treynor's ratio has the highest expected return (T^2 measure). To show this consider the portfolio p^*

$$R_{p^*} = aR_p + (1-a)R_f, \text{ with } a = 1/\beta_p. \tag{6.12}$$

This gives the mean and the beta of portfolio p^*

$$\mu^e_{p^*} = a\mu^e_p = \mu^e_p/\beta_p \tag{6.13}$$

$$\beta_{p^*} = a\beta_p = 1, \tag{6.14}$$

so the beta is one. We then define the T^2 measure as

$$T^2_p = \mu^e_{p^*} - \mu^e_m = \mu^e_p/\beta_p - \mu^e_m, \tag{6.15}$$

so the ranking (of fund p and q, say) in terms of Traynor's ratio and the T^2 are the same. See Example 6.2 and Figure 6.2 for an illustration.

The basic intuition is that with a *diversified portfolio* and *small investment*, idiosyncratic risk doesn't matter, only systematic risk (β) does. Compare with the setting of the Appraisal Ratio, where we also have a well diversified portfolio (the market), but the investment could be large.

Example 6.3 *(Additional portfolio risk) We hold a well diversified portfolio (d) and buy a fraction 0.05 of asset i (financed by borrowing), so the return is $R = R_d + 0.05\left(R_i - R_f\right)$. Suppose $\sigma^2_d = \sigma^2_i = 1$ and that the correlation of d and i is 0.25. The variance of R is then*

$$\sigma^2_d + \delta^2\sigma^2_i + 2\delta\sigma_{id} = 1 + 0.05^2 + 2 \times 0.05 \times 0.25 = 1 + 0.0025 + 0.025,$$

so the importance of the covariance is 10 times larger than the importance of the variance of asset i.

Proof. (*Version 1: Based on the beta representation.) The derivation of the beta representation shows that for all assets $\mu^e_i = \text{Cov}\left(R_i, R_m\right) A$, where A is some constant. Rearrange as $\mu^e_i/\beta_i = A\sigma^2_m$. A higher ratio than this is to be considered as a positive "abnormal" return and should prompt a higher investment. ■

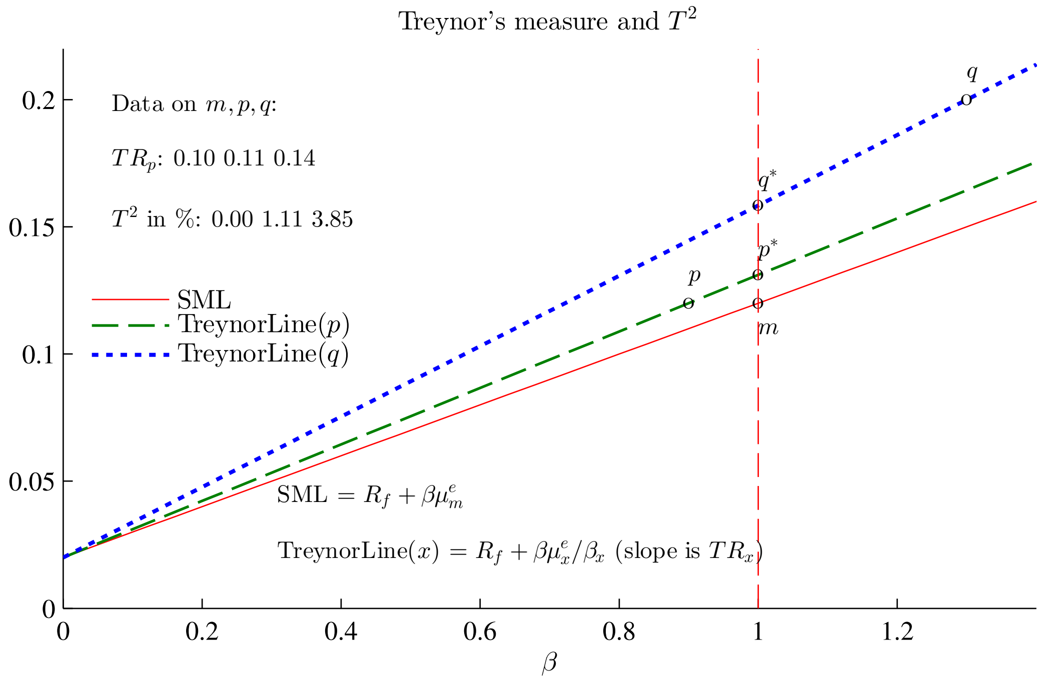


Figure 6.2: Treynor's ratio

Proof. (*Version 2: From first principles, kind of a proof...) Suppose we initially hold a well diversified portfolio (d) and we increase the position in asset i with the fraction δ by borrowing at the riskfree rate to get the return

$$R = R_d + \delta \left(R_i - R_f\right).$$

The incremental (compared to holding portfolio d) expected excess return is $\delta\mu_i^e$ and the incremental variance is $\delta^2\sigma_i^2 + 2\delta\sigma_{id} \approx 2\delta\sigma_{id}$, since δ^2 is very small. (The variance of R is $\sigma_d^2 + \delta^2\sigma_i^2 + 2\delta\sigma_{id}$.) To a first-order approximation, the change (E R_p − Var$(R_p)k/2$) in utility is therefore $\delta\mu_i^e - k\delta\sigma_{id}$, so a high value of μ_i^e/σ_{id} will increase utility. This suggests μ_i^e/σ_{id} as a performance measure. However, if portfolio d is indeed well diversified, then $\sigma_{id} \approx \sigma_{im}$. We could therefore use μ_i^e/σ_{im} or (by multiplying by σ_{mm}), μ_i^e/β_i as a performance measure. ■

6.1.5 Relationships among the Various Performance Measures

The different measures can give different answers when comparing portfolios, but they all share one thing: they are increasing in Jensen's alpha. By using the expected values from the CAPM regression ($\mu_p^e = \alpha_p + \beta_p \mu_m^e$), simple rearrangements give

$$\begin{aligned}
SR_p &= \frac{\alpha_p}{\sigma_p} + \text{Corr}(R_p, R_m) SR_m \\
AR_p &= \frac{\alpha_p}{\text{Std}(\varepsilon_{pt})} \\
TR_p &= \frac{\alpha_p}{\beta_p} + \mu_m^e.
\end{aligned} \tag{6.16}$$

and M^2 is just a scaling of the Sharpe ratio. Notice that these expressions do not assume that CAPM is the right pricing model—we just use the definition of the intercept and slope in the CAPM regression.

Since Jensen's alpha is the driving force in all these measurements, it is often used as performance measure in itself. In a sense, we are then studying how "mispriced" a fund is—compared to what it should be according to CAPM. That is, the alpha measures the "abnormal" return.

Proof. (of (6.16)*) Taking expectations of the CAPM regression (6.1) gives $\mu_p^e = \alpha_p + \beta_p \mu_m^e$, where $\beta_p = \text{Cov}(R_p, R_m)/\sigma_m^2$. The Sharpe ratio is therefore

$$SR_p = \frac{\mu_p^e}{\sigma_p} = \frac{\alpha_p}{\sigma_p} + \frac{\beta_p}{\sigma_p}\mu_m^e,$$

which can be written as in (6.16) since

$$\frac{\beta_p}{\sigma_p}\mu_m^e = \frac{\text{Cov}(R_p, R_m)}{\sigma_m \sigma_p}\frac{\mu_m^e}{\sigma_m}.$$

The AR_p in (6.16) is just a definition. The TR_p measure can be written

$$TR_p = \frac{\mu_p^e}{\beta_p} = \frac{\alpha_p}{\beta_p} + \mu_m^e,$$

where the second equality uses the expression for μ_p^e from above. ■

	α	SR	M^2	AR	Treynor	T^2
Market	0.000	0.318	0.000		6.250	0.000
Putnam	0.001	0.303	−0.296	0.000	6.251	0.001
Vanguard	2.495	0.505	3.679	0.548	10.765	4.515

Table 6.3: Performance Measures of Putnam Asset Allocation: Growth A and Vanguard Wellington, weekly data 1996:1-2013:5

6.1.6 Performance Measurement with More Sophisticated Benchmarks

Traditional performance tests typically rely on the alpha from a CAPM regression. The benchmark for the evaluation is then effectively a fixed portfolio consisting of assets that are correctly priced by the CAPM (obeys the beta representation). It often makes sense to use a more demanding benchmark. There are several popular alternatives.

If there are predictable movements in the market excess return, then it makes sense to add a "market timing" factor to the CAPM regression. For instance, Treynor and Mazuy (1966) argues that market timing is similar to having a beta that is linear in the market excess return

$$\beta_i = b_i + c_i R^e_{mt}. \tag{6.17}$$

Using in a traditional market model (CAPM) regression, $R^e_{it} = a_i + \beta_i R^e_{mt} + \varepsilon_{it}$, gives

$$R^e_{it} = a_i + b_i R^e_{mt} + c_i (R^e_{mt})^2 + \varepsilon_{it}, \tag{6.18}$$

where c captures the ability to "time" the market. That is, if the investor systematically gets out of the market (maybe investing in a riskfree asset) before low returns and vice versa, then the slope coefficient c is positive. The interpretation is not clear cut, however. If we still regard the market portfolio (or another fixed portfolio that obeys the beta representation) as the benchmark, then $a + c(R^e_{mt})^2$ should be counted as performance. In contrast, if we think that this sort of market timing is straightforward to implement, that is, if the benchmark is the market plus market timing, then only a should be counted as performance.

In other cases (especially when we think that CAPM gives systematic pricing errors), then the performance is measured by the intercept of a multifactor model like the Fama-French model.

A recent way to merge the ideas of market timing and multi-factor models is to allow the coefficients to be time-varying. In practice, the coefficients in period t are only allowed to be linear (or affine) functions of some information variables in an earlier period, z_{t-1}. To illustrate this, suppose z_{t-1} is a single variable, so the time-varying (or "conditional") CAPM regression is

$$\begin{aligned} R^e_{it} &= (a_i + \gamma_i z_{t-1}) + (b_i + \delta_i z_{t-1}) R^e_{mt} + \varepsilon_{it} \\ &= \theta_{i1} + \theta_{i2} z_{t-1} + \theta_{i3} R^e_{mt} + \theta_{i4} z_{t-1} R^e_{mt} + \varepsilon_{it}. \end{aligned} \tag{6.19}$$

Similar to the market timing regression, there are two possible interpretations of the results: if we still regard the market portfolio as the benchmark, then the other three terms should be counted as performance. In contrast, if the benchmark is a dynamic strategy in the market portfolio (where z_{t-1} is allowed to affect the choice market portfolio/riskfree asset), then only the first two terms are performance. In either case, the performance is time-varying.

6.2 Holdings-Based Performance Measurement

As a complement to the purely return-based performance measurements discussed, it may also be of interest to study how the portfolio weights change (if that information is available). This hightlights how the performance has been achieved.

Grinblatt and Titman's measure (in period t) is

$$GT_t = \sum\nolimits_{i=1}^{n} (w_{i,t-1} - w_{i,t-2}) R_{it}, \tag{6.20}$$

where $w_{i,t-1}$ is the weight on asset i in the portfolio chosen (at the end of) in period $t-1$ and $R_{i,t}$ is the return of that asset between (the end of) period $t-1$ and (end of) t. A positive value of GT_t indicates that the fund manager has moved into assets that turned out to give positive returns.

It is common to report a time-series average of GT_t, for instance over the sample $t = 1$ to T.

6.3 Performance Attribution

The performance of a fund is in many cases due to decisions taken on several levels. In order to get a better understanding of how the performance was generated, a performance attribution calculation can be very useful. It uses information on portfolio weights (for instance, in-house information) to decompose overall performance according to a number of criteria (typically related to different levels of decision making).

For instance, it could be to decompose the return (as a rough measure of the performance) into the effects of *(a)* allocation to asset classes (equities, bonds, bills); and *(b)* security choice within each asset class. Alternatively, for a pure equity portfolio, it could be the effects of *(a)* allocation to industries; and *(b)* security choice within each industry.

Consider portfolios p and b (for benchmark) from the same set of assets. Let n be the number of asset classes (or industries). Returns are

$$R_p = \sum_{i=1}^{n} w_i R_{pi} \text{ and } R_b = \sum_{i=1}^{n} v_i R_{bi}, \tag{6.21}$$

where w_i is the weight on asset class i (for instance, long T-bonds) in portfolio p, and v_i is the corresponding weight in the benchmark b. Analogously, R_{pi} is the return that the portfolio earns on asset class i, and R_{bi} is the return the benchmark earns. In practice, the benchmark returns are typically taken from well established indices.

Form the difference and rearrange to get

$$\begin{aligned} R_p - R_b &= \sum_{i=1}^{n} \left(w_i R_{pi} - v_i R_{bi}\right) \\ &= \underbrace{\sum_{i=1}^{n} (w_i - v_i) R_{bi}}_{\text{allocation effect}} + \underbrace{\sum_{i=1}^{n} w_i \left(R_{pi} - R_{bi}\right)}_{\text{selection effect}}. \end{aligned} \tag{6.22}$$

The first term is the *allocation effect* (that is, the importance of allocation across asset classes) and the second term is the *selection effect* (that is, the importance of selecting the individual securities within an asset class). In the first term, $(w_i - v_i) R_{bi}$ is the contribution from asset class (or industry) i. It uses the benchmark return for that asset class (as if you had invested in that index). Therefore the allocation effect simply measures the contribution from investing more/less in different asset class than the benchmark. If decisions on allocation to different asset classes are taken by senior management (or a board), then this is the contribution of that level. In the selection effect, $w_i \left(R_{pi} - R_{bi}\right)$ is the contribution of the security choice (within asset class i) since it measures the difference in returns (within that asset class) of the portfolio and the benchmark.

Remark 6.4 *(Alternative expression for the allocation effect*) The allocation effect is sometimes defined as $\sum_{i=1}^{n} (w_i - v_i)(R_{bi} - R_b)$, where R_b is the benchmark return. This is clearly the same as in (6.22) since $\sum_{i=1}^{n} (w_i - v_i) R_b = R_b \sum_{i=1}^{n} (w_i - v_i) = 0$ (as both sets of portfolio weights sum to unity).*

6.3.1 What Drives Differences in Performance across Funds?

Reference: Ibbotson and Kaplan (2000)

Plenty of research shows that the asset allocation (choice between markets or large market segments) is more important for mutual fund returns than the asset selection (choice of individual assets within a market segment). For other investors, including hedge funds, the leverage also plays a main role.

6.4 Style Analysis

Reference: Sharpe (1992)

Style analysis is a way to use econometric tools to find out the portfolio composition from a series of the returns, at least in broad terms.

The basic idea is to identify a number (5 to 10 perhaps) return indices that are expected to account for the brunt of the portfolio's returns, and then run a regression to find the portfolio "weights." It is essentially a multi-factor regression without any intercept and where the coefficients are constrained to sum to unity and to be positive

$$R_{pt}^{e} = \sum_{j=1}^{K} b_j R_{jt}^{e} + \varepsilon_{pt}, \text{ with} \tag{6.23}$$
$$\sum_{j=1}^{K} b_j = 1 \text{ and } b_j \geq 0 \text{ for all } j.$$

The coefficients are typically estimated by minimizing the sum of squared residuals. This is a nonlinear estimation problem, but there are very efficient methods for it (since it is a

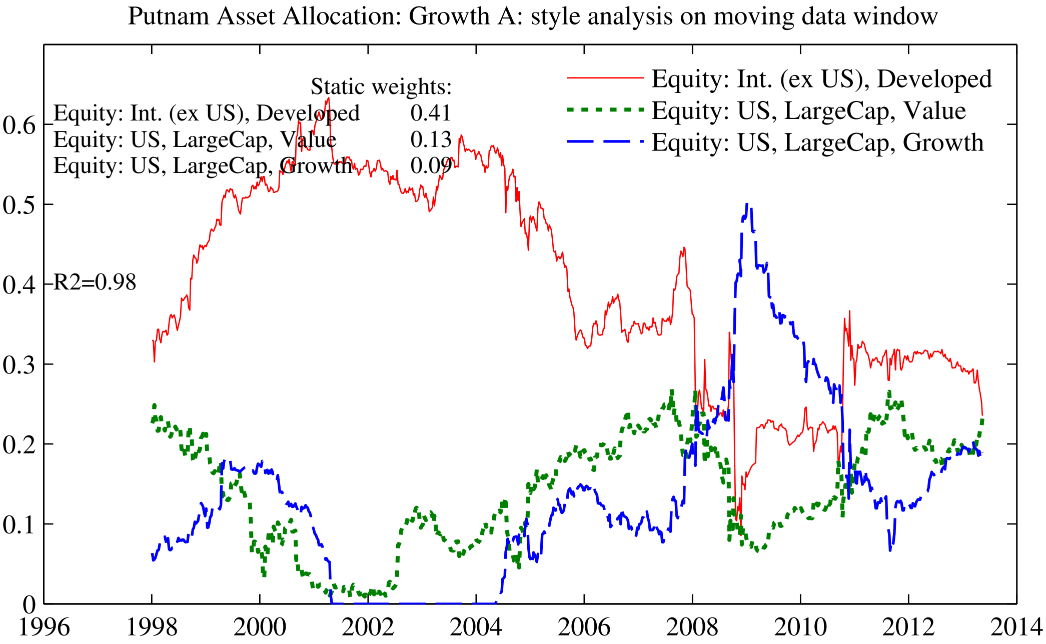


Figure 6.3: Example of style analysis, rolling data window

quadratic problem). Clearly, the restrictions could be changed to $U_j \leq b_j \leq L_j$, which could allow for short positions.

A pseudo-R^2 (the squared correlation of the fitted and actual values) is sometimes used to gauge how well the regression captures the returns of the portfolio. The residuals can be thought of as the effect of stock selection, or possibly changing portfolio weights more generally. One way to get a handle of the latter is to run the regression on a moving data sample. The time-varying weights are often compared with the returns on the indices to see if the weights were moved in the right direction.

See Figure 6.3 and Figure 6.5 for examples.

Bibliography

Elton, E. J., M. J. Gruber, S. J. Brown, and W. N. Goetzmann, 2010, *Modern portfolio theory and investment analysis*, John Wiley and Sons, 8th edn.

Ibbotson, R. G., and P. D. Kaplan, 2000, “Does Asset Allocation Policy Explain 40, 90 or

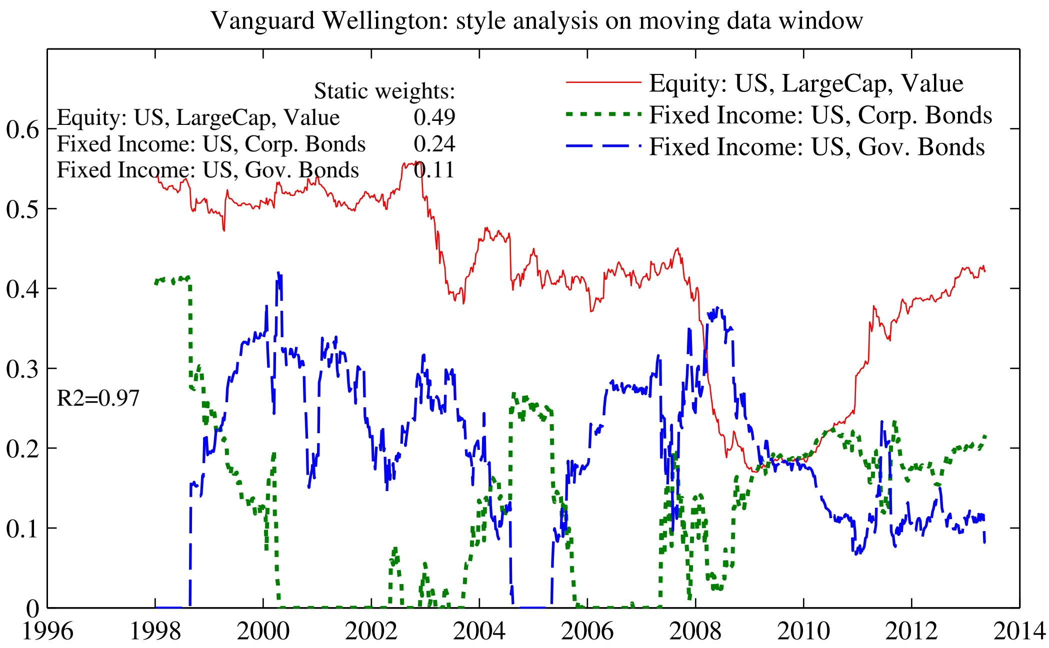


Figure 6.4: Example of style analysis, rolling data window

100 Percent of Performance?," *Financial Analysts Journal*, 65, 26–33.

Sharpe, W. F., 1992, "Asset allocation: management style and performance measurement," *Journal of Portfolio Management*, 39, 119–138.

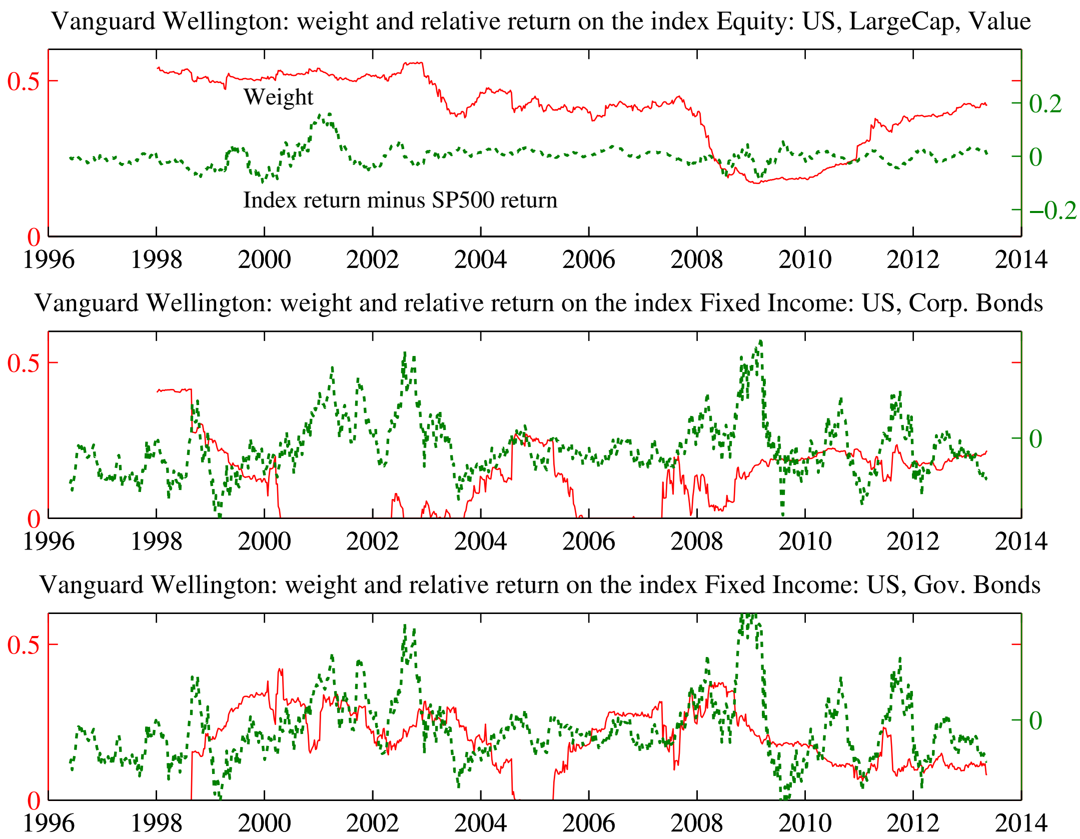


Figure 6.5: Style analysis and returns

7 Utility-Based Portfolio Choice

Reference: Elton, Gruber, Brown, and Goetzmann (2010) 12 and 18
Additional references: Danthine and Donaldson (2002) 5–6; Huang and Litzenberger (1988) 4–5; Cochrane (2001) 9 (5); Ingersoll (1987) 3–5 (6)

Material with a star (*) is not required reading.

7.1 Utility Functions and Risky Investments

Any model of portfolio choice must embody a notion of "what is best?" In finance, that often means a portfolio that strikes a good balance between expected return and its variance. However, in order to make sense of that idea—and to be able to go beyond it—we must go back to basic economic utility theory.

7.1.1 Specification of Utility Functions

In theoretical micro the utility function $U(x)$ is just an ordering without any meaning of the numerical values: $U(x) > U(y)$ only means that the bundle of goods x is preferred to y (but not by how much). In applied microeconomics we must typically be more specific than that by specifying the functional form of $U(x)$. As an example, to generate demand curves for two goods (x_1 and x_2), we may choose to specify the utility function as $U(x) = x_1^{\alpha} x_2^{1-a}$ (a Cobb-Douglas specification).

In finance (and quite a bit of microeconomics that incorporate uncertainty), the key features of the utility functions that we use are as follows.

First, utility is a function of a scalar argument, $U(x)$. This argument (x) can be end-of-period wealth, consumption or the portfolio return. In particular, we don't care about the composition of the consumption basket. In one-period investment problems, the choice of x is irrelevant since consumption equals wealth, which in turn is proportional to the portfolio return.

Second, uncertainty is incorporated by letting investors maximize expected utility, $\mathrm{E}\, U(x)$. Since returns (and therefore wealth and consumption) are uncertain, we need some way to rank portfolios at the time of investment (before the uncertainty has been resolved). In most cases, we use expected utility (see Section 7.1.2). As an example, suppose there are two states of the world: W (wealth) will be either 1 or 2 with probabilities $1/3$ and $2/3$. If $U(W) = \ln W$, then $\mathrm{E}\, U(W) = 1/3 \times \ln 1 + 2/3 \times \ln 2$.

Third, the functional form of the utility function is such that more is better and uncertainty is bad (investors are risk averse).

7.1.2 Expected Utility Theorem*

Expected utility, $\mathrm{E}\, U(W)$, is the right thing to maximize if the investors' preferences $U(W)$ are

1. complete: can rank all possible outcomes;
2. transitive: if A is better than B and B is better than C, then A is better than C (sounds like some basic form of consistency);
3. independent: if X and Y are equally preferred, and Z is some other outcome, then the following gambles are equally preferred

   $$X \text{ with prob } \pi \text{ and } Z \text{ with prob } 1 - \pi$$
   $$Y \text{ with prob } \pi \text{ and } Z \text{ with prob } 1 - \pi$$

   (this is the key assumption); and
4. such that every gamble has a certainty equivalent (a non-random outcome that gives the same utility, fairly trivial).

7.1.3 Basic Properties of Utility Functions: (1) More is Better

The idea that *more is better* (nonsatiation) is almost trivial. If $U(W)$ is differentiable, then this is the same as that marginal utility is positive, $U'(W) > 0$.

Example 7.1 *(Logarithmic utility)* $U(W) = \ln W$ *so* $U'(W) = 1/W$ *(assuming* $W > 0$*).*

7.1.4 Basic Properties of Utility Functions: (2) Risk is Bad

With a utility function, *risk aversion* (uncertainty is considered to be bad) is captured by the concavity of the function.

As an example, consider Figure 7.1. It shows a case where the portfolio (or wealth, or consumption,...) of an investor will be worth Z_- or Z_+, each with a probability of a half. This utility function shows risk aversion since the utility of getting the expected payoff for sure is higher than the expected utility from owning the uncertain asset

$$U(\mathrm{E}\,Z) > 0.5U(Z_-) + 0.5U(Z_+) = \mathrm{E}\,U(Z). \tag{7.1}$$

This is a way of saying that the investor does not like risk.

Rearranging gives

$$U(\mathrm{E}\,Z) - U(Z_-) > U(Z_+) - U(\mathrm{E}\,Z), \tag{7.2}$$

which says that a loss (left hand side) counts for more than a gain of the same amount. Another way to phrase the same thing is that a poor person appreciates an extra dollar more than a rich person. This is a key property of a concave utility function—and it has an immediate effect on risk premia.

The (lowest) price (P) the investor is willing to sell this portfolio for is the certain amount of money which gives the same utility as $\mathrm{E}\,U\,(Z)$, that is, the value of P that solves the equation

$$U(P) = \mathrm{E}\,U(Z). \tag{7.3}$$

This price P is also called the *certainty equivalent* of the portfolio. From (7.1) we know that this utility is lower than the utility from the expected payoff, $U(P) < U(\mathrm{E}\,Z)$. We also know that the utility function is an increasing function. It then follows directly that the price is lower than the expected payoff

$$P < \mathrm{E}\,Z = 0.5Z_- + 0.5Z_+. \tag{7.4}$$

See Figures 7.1–7.2 for an illustration.

Example 7.2 *(Certainty equivalent) Suppose you have a CRRA utility function and own an asset that gives either 85 or 115 with equal probability. What is the certainty equivalent*

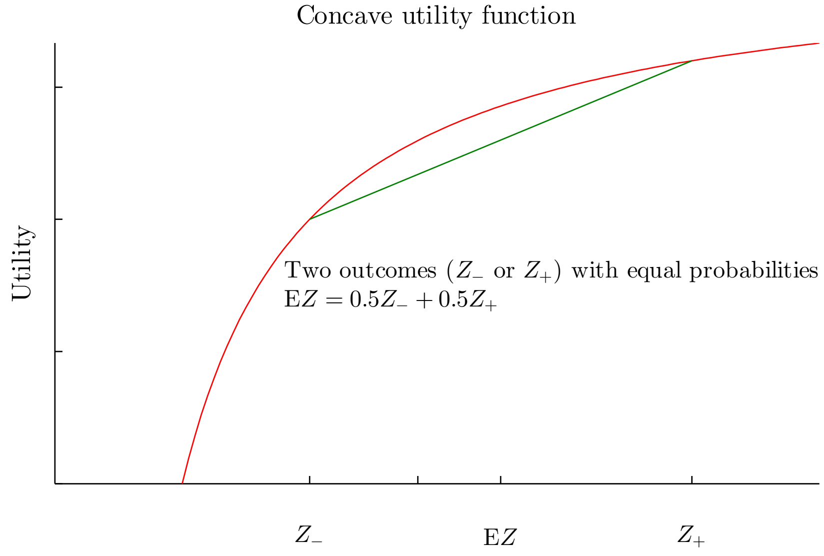


Figure 7.1: Utility function

(that is, the lowest price you would sell this asset for)? The answer is the P that solves

$$\frac{P^{1-k}}{1-k} = 0.5\frac{85^{1-k}}{1-k} + 0.5\frac{115^{1-k}}{1-k}.$$

(The answer is $P = (0.5 \times 85^{1-k} + 0.5 \times 115^{1-k})^{1/(1-k)}$.) For instance, with $k = 0$, 2, 5, 10, and 25 we have $P \approx 100$, 97.75, 94.69, 91.16, and 87.49. Note that if we scale the asset payoffs (here 85 and 115) with some factor, then the price is scaled with the same factor. This is a typical feature of the CRRA utility function.

This means that the expected net return on the risky portfolio that the investor demands is

$$\mathrm{E}\, R_Z = \frac{\mathrm{E}\, Z}{P} - 1 > 0, \tag{7.5}$$

which is greater than zero. This "required return" is higher if the investor is very risk averse (very concave utility function). On the other hand, it goes towards zero as the investor becomes less and less risk averse (the utility function becomes more and more linear). In the limit (a risk neutral investor), the required return is zero. Loosely speaking, we can think of $\mathrm{E}\, R_Z$ as a *risk premium* (more generally, the risk premium is $\mathrm{E}\, R_Z$ minus a riskfree rate). Notice that this analysis applies to the portfolio (or wealth, or consumption,...) that is the argument of the utility function—not to any individual asset. To

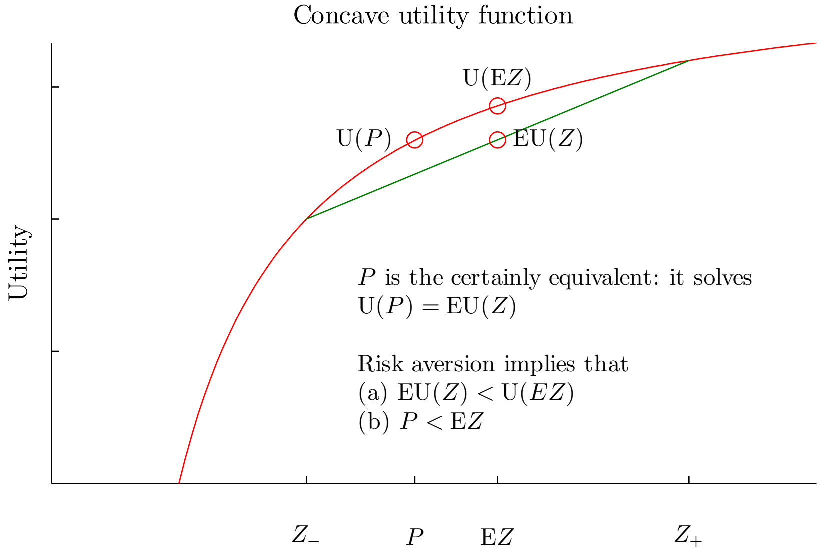


Figure 7.2: Certainty equivalent

analyse an individual asset, we need to study how it changes the argument of the utility function, so the covariances with the other assets play a key role.

Example 7.3 *(Utility and two states) Suppose the utility function is logarithmic and that* $(Z_-, Z_+) = (1, 2)$*. Then, expected utility in (7.1) is*

$$\mathrm{E}\, U\,(Z) = 0.5 \ln 1 + 0.5 \ln 2 \approx 0.35,$$

so the price must be such that

$$\ln P \approx 0.35, \textit{ that is, } P \approx e^{0.35} \approx 1.41.$$

The expected return (7.5) is

$$(0.5 \times 1 + 0.5 \times 2)\,/1.41 \approx 1.06.$$

7.1.5 Is Risk Aversion Related to the Level of Wealth?

We now take a closer look at what the functional form of the utility function implies for investment choices. In particular, we study if risk aversion will be related to the wealth level.

First, define *absolute risk aversion* as

$$A(W) = \frac{-U''(W)}{U'(W)}, \tag{7.6}$$

where $U'(W)$ is the first derivative and $U''(W)$ the second derivative. Second, define *relative risk aversion* as

$$R(W) = WA(W) = \frac{-WU''(W)}{U'(W)}. \tag{7.7}$$

These two definitions are strongly related to the attitude towards taking risk.

Consider an investor with wealth W who can choose between taking on a zero mean risk Z (so $\mathrm{E}\,Z = 0$) or pay a price P. He is indifferent if

$$\mathrm{E}\,U(W + Z) = U(W - P). \tag{7.8}$$

If Z is a small risk, then we can make a second order approximation

$$P \approx A(W)\,\mathrm{Var}(Z)/2, \tag{7.9}$$

which says that the price the investor is willing to pay to avoid the risk Z is proportional to the absolute risk aversion $A(W)$.

Proof. (of (7.9)) Approximate as

$$\begin{aligned} \mathrm{E}\,U(W + Z) &\approx U(W) + U'(W)\,\mathrm{E}\,Z + U''(W)\,\mathrm{E}\,Z^2/2 \\ &= U(W) + U''(W)\,\mathrm{Var}(Z)/2, \end{aligned}$$

since $\mathrm{E}\,Z = 0$. (We here follow the rule of adding terms to the Taylor approximation to have two left after taking expectations.) Now, approximate $U(W - P) \approx U(W) - U'(W)P$. Set equal to get (7.9). ■

If we change the example in (7.8)–(7.9) to make the risk proportional to wealth, that is $Z = Wz$ where z is the risk factor, then (7.9) directly gives

$$\begin{aligned} P &\approx A(W)W^2\,\mathrm{Var}(z)/2, \text{ so} \\ P/W &\approx R(W)\,\mathrm{Var}(z)/2, \end{aligned} \tag{7.10}$$

which says that the fraction of wealth (P/W) that the investor is willing to pay to avoid

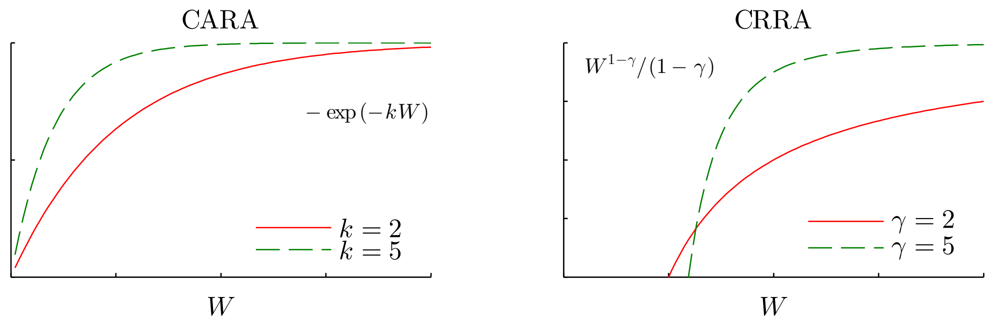


Figure 7.3: Examples of utility functions

the risk (z) is proportional to the relative risk aversion $R(W)$.

These results mostly carry over to the portfolio choice: high absolute risk aversion typically implies that only small *amounts* are invested into risky assets, whereas a high relative risk aversion typically leads to small *portfolio weights* of risky assets.

Figure 7.3 demonstrates a number of commonly used utility functions, and the following discussion outlines their main properties.

Remark 7.4 *(Mean-variance utility and portfolio choice) Suppose expected utility is* $\mathrm{E}(1+R_p)W_0 - k\,\mathrm{Var}[(1+R_p)W_0]/2$ *where* W_0 *is initial wealth and the portfolio return is* $R_p = vR_1 + (1-v)R_f$*, where* R_1 *is a risky asset and* R_f *a riskfree asset. The optimal portfolio weight is*

$$v = \frac{1}{kW_0}\frac{\mathrm{E}\,R_1 - R_f}{\mathrm{Var}(R_1)}.$$

A poor investor therefore invests the same amount in the risky asset as a rich investor (vW_0 does not depend on W_0), and his portfolio weight on the risky asset (v) is larger.

The *CARA utility function* (constant absolute risk aversion), $U(W) = -e^{-kW}$, is also quite simple to use (in particular when returns are normally distributed—see below), but has the unappealing feature that the amount invested in the risky asset (in a risky/riskfree trade-off) is constant across (initial) wealth levels. This means, of course, that wealthy investors have a lower portfolio weight on risky assets.

Remark 7.5 *(Risk aversion in CARA utility function)* $U(W) = -e^{-kW}$ *gives* $U'(W) = ke^{-kW}$ *and* $U''(W) = -k^2e^{-kW}$*, so we have* $A(W) = k$*. This means an increasing* *relative risk aversion, $R(W) = Wk$, so a poor investor typically has a larger portfolio weight on the risky asset than a rich investor.*

The *CRRA utility function* (constant relative risk aversion) is often harder to work with, but has the nice property that the portfolio weights are unaffected by the initial wealth (once again, see the following remark for the algebra). Most evidence suggests that the CRRA utility function fits data best. For instance, historical data show no trends in portfolio weights or risk premia—in spite of investors having become much richer over time.

Remark 7.6 *(Risk aversion in CRRA utility function) $U(W) = W^{1-k}/(1-k)$ gives $U'(W) = W^{-k}$ and $U''(W) = -kW^{-k-1}$, so we have $A(W) = k/W$ and $R(W) = k$. The absolute risk aversion decreases with the wealth level in such a way that the relative risk aversion is constant. In this case, a poor investor typically has the same portfolio weight on the risky asset as a rich investor.*

7.2 Utility-Based Portfolio Choice and Mean-Variance Frontiers

7.2.1 Utility-Based Portfolio Choice

Suppose the investor maximizes expected utility from wealth by choosing between a risky and a riskfree asset

$$\max_v \mathrm{E}\, U(R_p), \text{ with } R_p = vR_1^e + R_f. \tag{7.11}$$

The first order condition with respect to the weight on risky assets is

$$0 = \frac{\partial \mathrm{E}\, U(vR_1^e + R_f)}{\partial v} = \mathrm{E}[U'(vR_1^e + R_f) \times R_1^e], \tag{7.12}$$

where $U'(vR_1^e + R_f)$ is shorthand notation for the marginal utility, evaluated at $vR_1^e + R_f$. Notice that the expectation on the RHS is the expectation of the product of marginal utility and the excess return. Also notice that the order of E and ∂ are different on the LHS and RHS. This is permissable since E defines a sum (and a derivative of a sum is the sum of derivatives, see below for a remark).

Remark 7.7 *(Interchanging the order of* E *and ∂*) Recall that for two functions $f(x)$ and $g(x)$ we have*

$$\frac{\partial}{\partial v}[f(x) + g(x)] = \frac{\partial f(x)}{\partial v} + \frac{\partial g(x)}{\partial v}.$$

That is, a derivative of a sum equals the the sum of a derivatives. We can apply this by supposing that R^e_1 can take on S different values (denoted $R^e_{1,s}$), with the probabilities π_s. We can then write $U(R_p) = \sum_{s=1}^{S} \pi_s U(vR^e_{1,s} + R_f)$. Differentiating expected utility gives

$$\frac{\partial \operatorname{E} U(R_p)}{\partial v} = \frac{\partial}{\partial v} \sum_{s=1}^{S} \pi_s U(vR^e_{1,s} + R_f) = \sum_{s=1}^{S} \pi_s \frac{\partial U(vR^e_{1,s} + R_f)}{\partial v} = \operatorname{E} \frac{\partial U(R_p)}{\partial v}.$$

This shows that $\partial \operatorname{E} U(R_p)/\partial v = \operatorname{E}[\partial U(R_p)/\partial v]$.

Clearly, the first order condition (7.12) defines one equation in one unknown (v). Suppose we have chosen some utility function and that we know the distribution of the returns—it should then be possible to solve for the portfolio weight. Unfortunately, that can be fairly complicated. For instance, utility might be highly non-linear so the calculation of its expected value involves difficult integrations (possibly requiring numerical methods since there is no analytical solution). With many assets there are many first order conditions, so the system of equations can be large.

Example 7.8 *(Portfolio choice with log utility and two states) Suppose $U(R_p) = \ln R_p$, and that there is only one risky asset. The excess return on the risky asset R^e is either a low value R^{e-} (with probability π) or a high value R^{e+}(with probability $1 - \pi$). The optimization problem is then*

$$\max_v \operatorname{E} U\left(R_p\right) \text{ where } \operatorname{E} U\left(R_p\right) = \pi \ln\left(vR^{e-} + R_f\right) + (1 - \pi) \ln\left(vR^{e+} + R_f\right).$$

The first order condition ($\partial \operatorname{E} U\left(R_p\right)/\partial v = 0$) is

$$\pi \frac{R^{e-}}{vR^{e-} + R_f} + (1 - \pi) \frac{R^{e+}}{vR^{e+} + R_f} = 0,$$

so we can solve for the portfolio weight as

$$v = -R_f \frac{\pi R^{e-} + (1 - \pi) R^{e+}}{R^{e-}R^{e+}}.$$

For instance, with $R_f = 1.1$, $R^{e-} = -0.3$, $R^{e+} = 0.4$, and $\pi = 0.5$, we get

$$v = -1.1 \frac{0.5 \times (-0.3) + (1 - 0.5)\, 0.4}{(-0.3) \times 0.4} \approx 0.46.$$

See Figure 7.4 for an illustration.

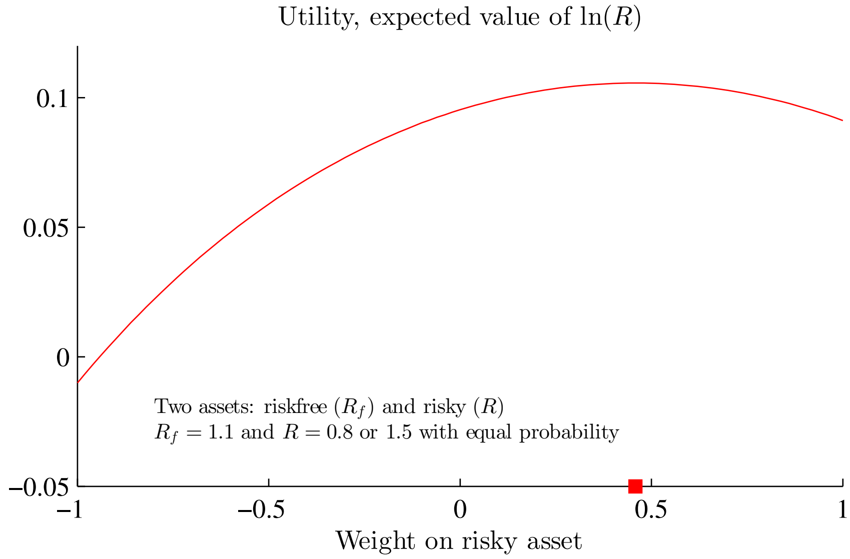


Figure 7.4: Example of portfolio choice with a log utility function

Suppose $v = 0$ (no investment in the risky asset) would be an optimal decision, then the portfolio return equals the riskfree rate which is not random. The expression on the right hand side of the first order condition (7.12) can then be written

$$\begin{aligned} \mathrm{E}[U'(R_f)R_1^e] = U'(R_f)\,\mathrm{E}\,R_1^e = 0 \text{ if} \\ \mathrm{E}\,R_1^e = 0. \end{aligned} \tag{7.13}$$

This shows that *no investment in the risky asset is optimal when its expected excess return is zero*. (Why take on risk if it does not give any benefits?) In contrast, if $\mathrm{E}\,R_1^e > 0$, then $v = 0$ cannot be optimal.

7.2.2 General Utility-Based Portfolio Choice

For simplicity, assume that consumption equals wealth, which we normalize to unity. The optimization problem with a general utility function, n risky and a riskfree asset is then

$$\max_{v_1, v_2, \ldots} \mathrm{E}\,U\left(R_p\right), \text{ where} \tag{7.14}$$

$$R_p = \textstyle\sum_{i=1}^{n} v_i R_i^e + R_f. \tag{7.15}$$

where R_i^e is the excess return on asset i and R_f is a riskfree rate. The first order conditions for the portfolio weights are

$$\frac{\partial \operatorname{E} U(R_p)}{\partial v_i} = 0 \text{ for } i = 1, 2, ..., n \tag{7.16}$$

which defines n equations in n unknowns: $v_1, v_2, ..., v_n$. As discussed before, the explicit solution is often hard to obtain—so it would be convenient if we could simplify the problem.

7.2.3 Is the Optimal Portfolio on the Mean-Variance Frontier?

There are important cases where we can side-step most of the problems with solving (7.16)—since it can be shown that the portfolio choice will actually be such that a portfolio on the minimum-variance frontier (upper MV frontier) will be chosen.

The optimal portfolio must be on the minimum-variance frontier when expected utility can be (re-)written as a function in terms of the expected return (increasing) and the variance (decreasing) only, that is

$$\operatorname{E} U\left(R_p\right) = V(\mu_p, \sigma_p^2), \tag{7.17}$$
$$\text{with } \partial V(\mu_p, \sigma_p^2)/\partial \mu_p > 0 \text{ and } \partial V(\mu_p, \sigma_p^2)/\partial \sigma_p^2 < 0.$$

For an illustration, see Figure 7.5 which shows the isoutility curves (curves with equal utility) from a mean-variance utility function ($\operatorname{E} U(R_p) = \mu_p - (k/2)\,\sigma_p^2$). Whenever expected utility obeys (7.17) (not just for the mean-variance utility function) the isoutility curves will look similar—so the optimum is on the minimum-variance frontier. The intuition behind (7.17) is that an investor wants to move as far to the north-west as possible in Figure 7.5—but that he/she is willing to trade off lower expected returns for lower volatility, that is, has isoutility functions as in the figure. What is possible is clearly given by the mean-variance frontier—so the solution is a point on the upper frontier. (This can also be shown algebraically, but it is slightly messy.) Conditions for (7.17) are discussed below.

In the case with both a riskfree and risky assets, this means that all investors (provided they have the same beliefs) will pick some mix of the riskfree asset and the *tangency portfolio* (where the ray from the riskfree rate is tangent to the mean-variance frontier of risky assets). This is the *two-fund theorem*. Notice that all this says is that the optimal

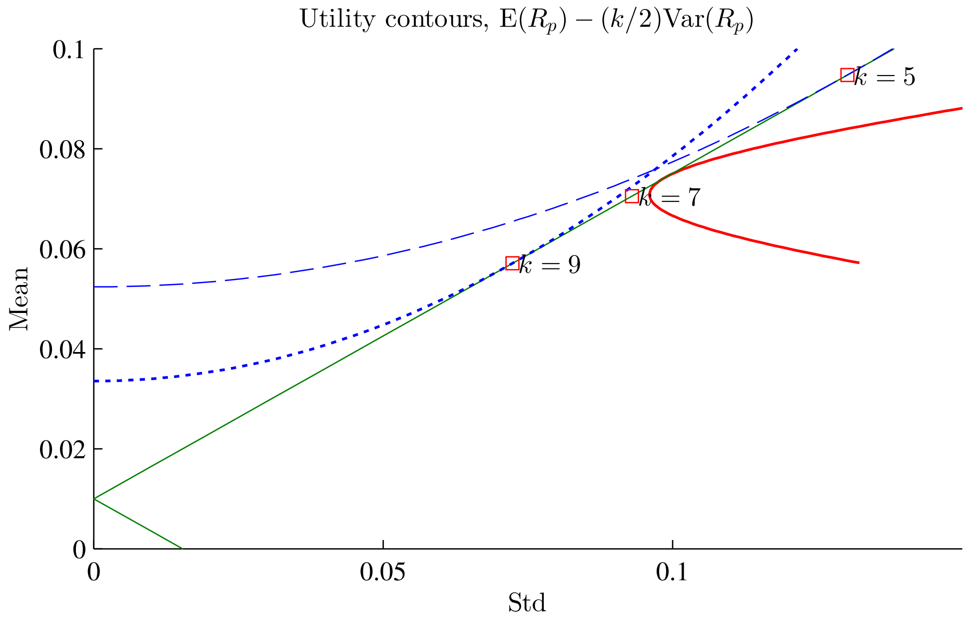


Figure 7.5: Iso-utility curves, mean-variance utility with different risk aversions

portfolio is *somewhere on the mean variance frontier*. We cannot tell exactly where unless we are more precise about the exact form of the preferences.

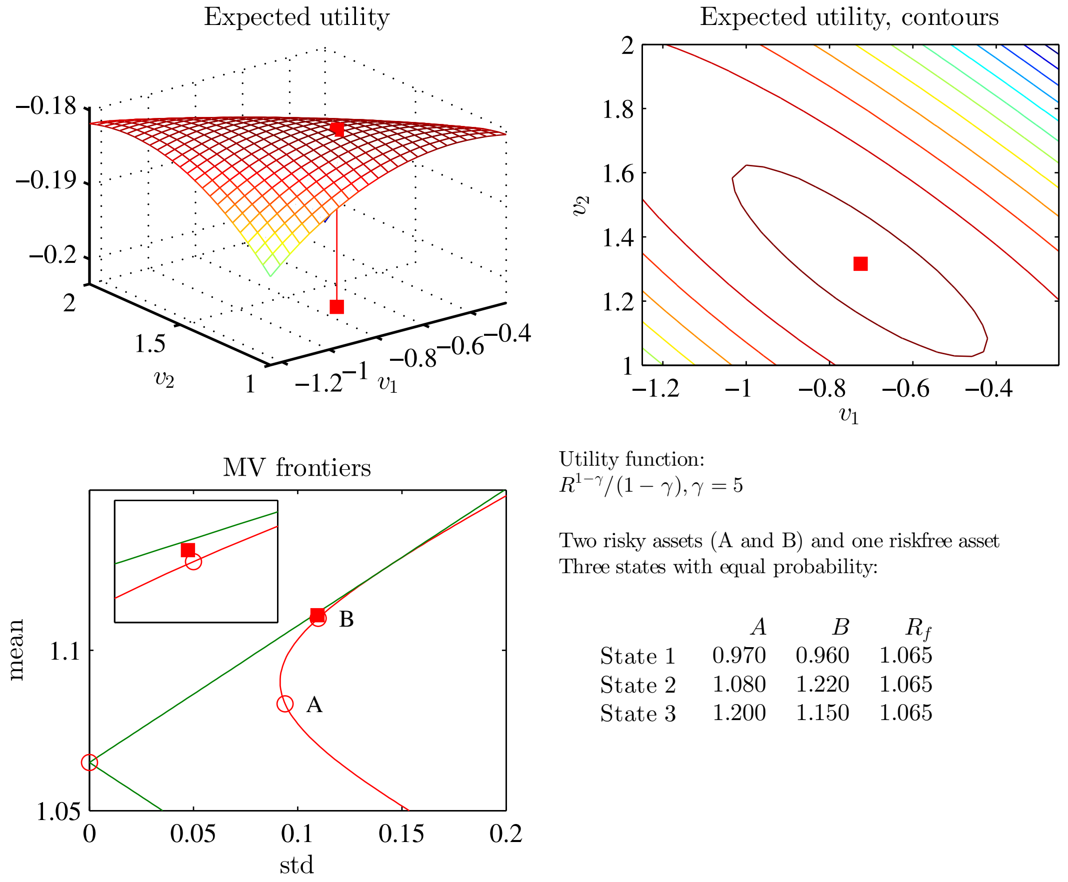
See Figures 7.6–7.7 for examples of cases when we do not get a mean-variance portfolio.

7.2.4 Special Cases

This section outlines special cases when the utility-based portfolio choice problem can be rewritten as in (7.17) (in terms of mean and variance only), so that the optimal portfolio belongs to the minimum-variance set. (Recall that with a riskfree asset this minimum-variance set is a ray that starts at R_f and goes through the tangency portfolio.)

Case 1: Mean-Variance Utility

We know that if the investor maximizes $\mathrm{E}\, R_p - \mathrm{Var}(R_p)k/2$, then the optimal portfolio is on the mean-variance frontier. Clearly, this is the same as assuming that the utility function is $U(R_p) = R_p - (R_p - \mathrm{E}\, R_p)^2 k/2$ (evaluate $\mathrm{E}\, U(R_p)$ to see this).



	A	B	R_f
State 1	0.970	0.960	1.065
State 2	1.080	1.220	1.065
State 3	1.200	1.150	1.065

Figure 7.6: Example of when the optimal portfolio is (very slightly) off the MV frontier

Case 2: Quadratic Utility

If utility is quadratic in the return (or equivalently, in wealth)

$$U(R_p) = R_p - bR_p^2/2, \tag{7.18}$$

then expected utility can be written

$$\begin{aligned} \mathrm{E}\, U(R_p) &= \mathrm{E}\, R_p - b\, \mathrm{E}\, R_p^2/2 \\ &= \mathrm{E}\, R_p - b[\mathrm{Var}(R_p) + (\mathrm{E}\, R_p)^2]/2 \end{aligned} \tag{7.19}$$

since $\mathrm{Var}(R_p) = \mathrm{E}\, R_p^2 - (\mathrm{E}\, R_p)^2$. (We assume that all these moments are finite.) For $b > 0$ this function is decreasing in the variance, and increasing in the mean return (as

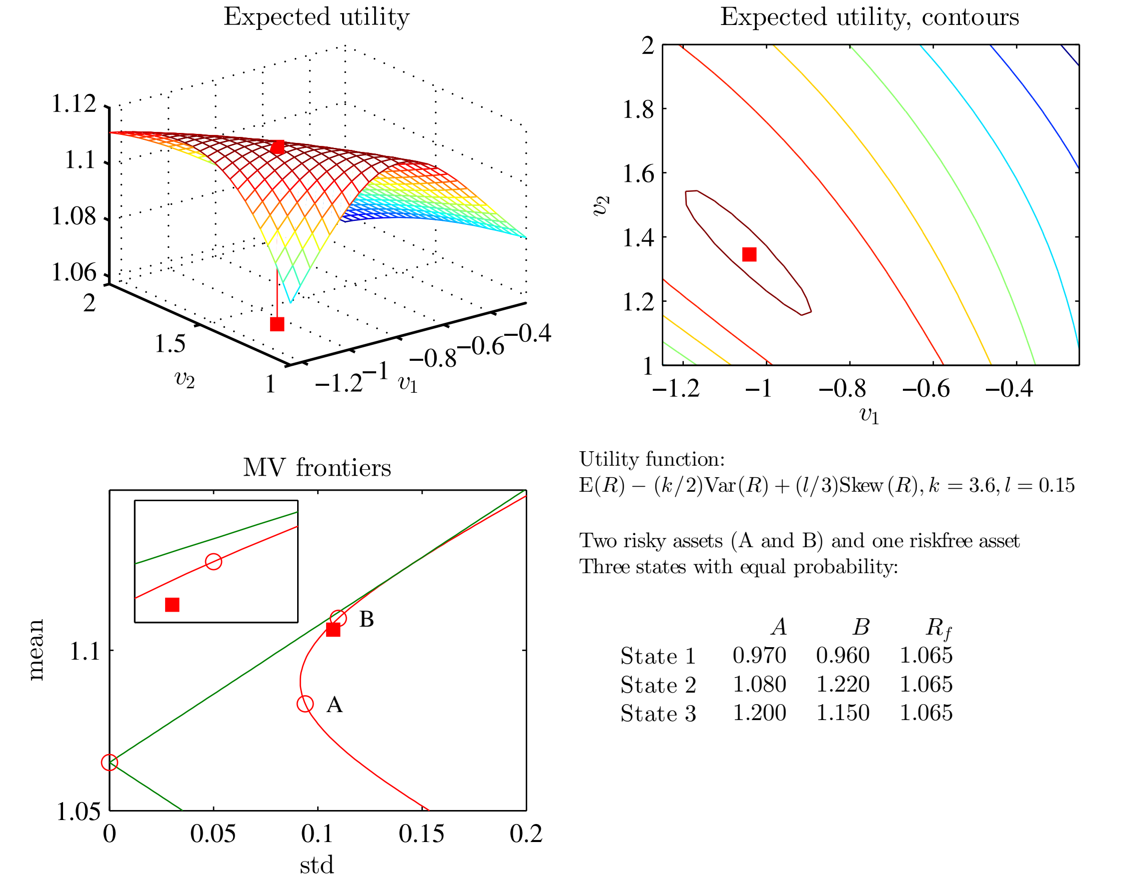


Figure 7.7: Example of when the optimal portfolio is (very slightly) off the MV frontier

long as $b \operatorname{E} R_p < 1$). The optimal portfolio is therefore on the minimum-variance frontier. See Figure 7.9 for an example.

The main drawback with this utility function is that we have to make sure that we are on the portion of the curve where utility is increasing (below the so called "bliss point"). Moreover, the quadratic utility function has the strange property that the amount invested in risky assets decreases as wealth increases (increasing absolute risk aversion).

Case 3: Normally Distributed Returns

When the distribution of any *portfolio* return is fully described by the mean and variance, then maximizing $\operatorname{E} U(R_p)$ will result in a mean variance portfolio—under some extra assumptions about the utility function discussed below. A normal distribution (among a few other distributions) is completely described by its mean and variance. Moreover, any portfolio return would be normally distributed if the returns on the individual assets have a multivariate normal distribution (recall: $x + y$ is normally distributed if x and y are).

The extra assumptions needed are that utility is strictly increasing in wealth ($U'(R_p) > 0$), displays risk aversion ($U''(R_p) < 0$), and utility must be defined for all possible outcomes. The later sounds trivial, but it is not. For instance, the logarithmic utility function $U(R_p) = \ln R_p$ cannot be combined with returns (end of period wealth) that can take negative values (for instance, $\ln(-1) = \pi i$ which is not a real number which is something we require from a utility function).

Remark 7.9 *(Taylor series expansion) Recall that a Taylor series expansion of a function $f(x)$ around the point x_0 is $f(x) = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{d^n f(x_0)}{dx^n}(x - x_0)^n$, where $d^n f(x_0)/dx^n$ is the nth derivative of $f()$ evaluated at x_0 and $n!$ is the factorial ($n! = 1 \times 2 \times \ldots \times n$ and $0! = 1$ by definition).*

Do a Taylor series expansion of the utility function $U(R_p)$ around the average portfolio return ($\mathrm{E}\, R_p$) to get

$$U(R_p) = \sum\nolimits_{n=0}^{\infty} \frac{1}{n!} \frac{d^n U(\mathrm{E}\, R_p)}{dW^n} \left(R_p - \mathrm{E}\, R_p\right)^n, \tag{7.20}$$

where $d^n U(\mathrm{E}\, R_p)/dW^n$ denotes the nth derivative of the utility function—evaluated at the point $\mathrm{E}\, R_p$.

Take expectations, notice that the randomness is only in the $\left(R_p - \mathrm{E}\, R_p\right)^n$ terms and recall that $\mathrm{E}\left(R_p - \mathrm{E}\, R_p\right) = 0$ and that $\mathrm{E}\left(R_p - \mathrm{E}\, R_p\right)^2 = \mathrm{Var}(R_p)$. (As usual, $\mathrm{E}\left(R_p - \mathrm{E}\, R_p\right)^2$ should be understood as $\mathrm{E}[\left(R_p - \mathrm{E}\, R_p\right)^2]$.) Write out as

$$\mathrm{E}\, U(R_p) = U(\mathrm{E}\, R_p) + 0 + \frac{1}{2} U''(\mathrm{E}\, R_p)\, \mathrm{Var}(R_p) + \sum\nolimits_{n=3}^{\infty} \frac{1}{n!} \frac{d^n U(\mathrm{E}\, R_p)}{dW^n} \mathrm{E}\left(R_p - \mathrm{E}\, R_p\right)^n. \tag{7.21}$$

Remark 7.10 *(Taylor expansion of a CRRA utility function*) For a CRRA utility function, $(1 + R_p)^{1-\gamma}/(1 - \gamma)$, we have*

$$U''(\mathrm{E}\, R_p) = -\gamma(1 + \mathrm{E}\, R_p)^{-\gamma-1} < 0 \text{ and } U'''(\mathrm{E}\, R_p) = \gamma(1 + \gamma)(1 + \mu_p)^{-\gamma-2} > 0,$$

so variance is bad, but skewness is good.

Remark 7.11 *(Higher central moments for a normal distribution) If x is normally distributed, then* $\mathrm{E}\,(x-\mu)^n = 0$ *if n is odd and proportional to* $\mathrm{Var}(x)$ *if n is even. To be precise, for even n,* $\mathrm{E}\,(x-\mu)^n = \mathrm{Var}(x) \times (n-1)!!$, *where* $(n-1)!!$ *is the product of all odd numbers up to and including $n-1$,* $1 \times 3 \times \ldots \times (n-3) \times (n-1)$.

If R_p is normally distributed, then $\mathrm{E}\left(R_p - \mathrm{E}\,R_p\right)^n = 0$ if n is odd and proportional to $\mathrm{Var}(R_p)$ if n is even. This means that (7.21) can be written

$$\mathrm{E}\,U(R_p) = U(\mathrm{E}\,R_p) + F(\mathrm{E}\,R_p)\,\mathrm{Var}(R_p), \tag{7.22}$$

where F is a (complicated) function of the mean return. The idea is essentially that the mean and variance fully describe the normal distribution. Since increasing concave utility functions are increasing in the mean and decreasing in the variance (of the portfolio return), the result is quite intuitive.

Normally distributed returns should be considered as an approximation for three reason. First, limited liability means that the gross return can never be negative (the asset price cannot be negative), that is, the simple net return can never be less than -100%. A normal distribution cannot rule out this possibility (although it may have a very low probability). Second, option returns have distributions which are clearly different from normal distributions: a lot of probability mass at exactly -100% (no exercise) and then a continuous distribution for higher returns. Third, empirical evidence suggests that most asset returns have distributions with fatter tails and more skewness than implied by a normal distribution, especially when the returns are measured over short horizons.

As an illustration, suppose the investor maximizes a utility function with *constant absolute risk aversion* $k > 0$

$$U(R_p) = -\exp(-R_p k). \tag{7.23}$$

(It is straightforward to show that this utility function satisfies the extra conditions.)

Proposition 7.12 *If returns are normally distributed, then maximizing the expected value of the CARA utility function is the same as solving a mean-variance problem.*

Proof. (of Proposition 7.12) First, recall that if $x \sim N\left(\mu, \sigma^2\right)$, then $\mathrm{E}\,e^x = e^{\mu+\sigma^2/2}$. Therefore, rewrite expected utility as

$$\mathrm{E}\,U(R_p) = \mathrm{E}\left[-\exp\left(-R_p k\right)\right] = -\exp\left[-\mathrm{E}\,R_p k + \mathrm{Var}(R_p)k^2/2\right].$$

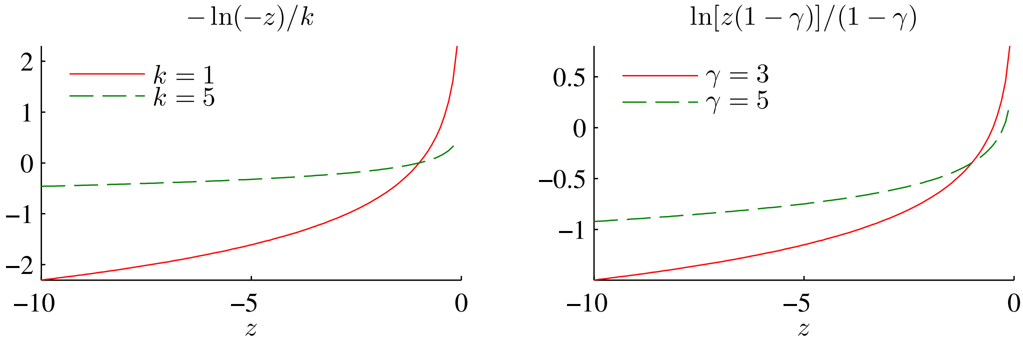


Figure 7.8: Transforming expected utility

Notice that the assumption of normally distributed returns is crucial for this result. Second, recall that if x maximizes (minimizes) $f(x)$, then it also maximizes (minimizes) $g[f(x)]$ if g is a strictly increasing function. The function $-\ln(-z)/k$ is defined for $z < 0$ and it is increasing in z, see Figure 7.8. We can apply this function by letting z be the right hand side of the previous equation to get

$$-\ln(-z)/k = \mathrm{E}\, R_p - \mathrm{Var}(R_p)k/2.$$

Therefore, maximizing the expected CARA utility or MV preferences (in terms of the returns) gives the same solution. (When utility is written in terms of wealth $W_0(1 + R_p)$ where R_p is the portfolio return, the last equation becomes $W_0\, \mathrm{E}(1+R_p) - W_0^2\, \mathrm{Var}(R_p)k/2$.)
■

Case 4: CRRA Utility and Lognormally Distributed Portfolio Returns

Proposition 7.13 *Consider a CRRA utility function,* $(1 + R_p)^{1-\gamma}/(1-\gamma)$*, and suppose all log portfolio returns,* $r_p = \ln(1+R_p)$*, happen to be normally distributed. The solution is then, once again, on the mean-variance frontier.*

This result is especially useful in analysis of multi-period investments. (Notice, however, that this should be thought of as an approximation since $1 + R_p = \alpha(1 + R_1) + (1-\alpha)(1 + R_2)$ is not lognormally distributed even if both R_1 and R_2 are.)

See Figure 7.9 for an example.

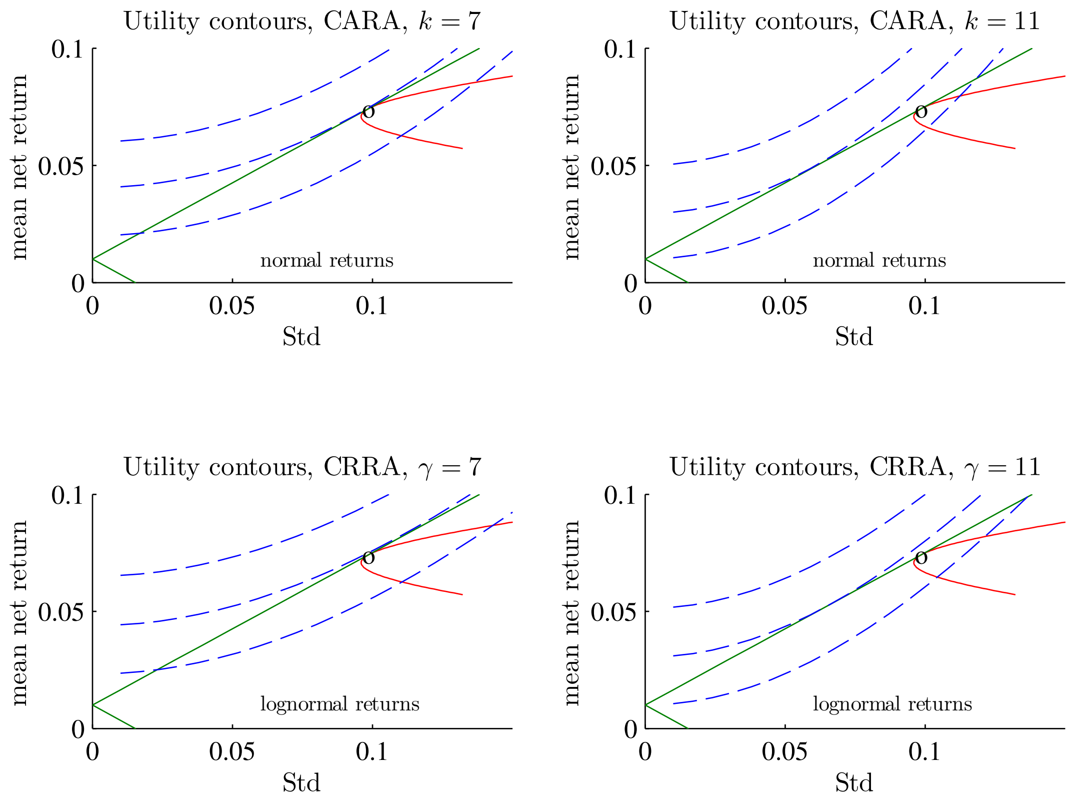


Figure 7.9: Contours with same utility level when returns are normally or lognormally distributed. The means and standard deviations (on the axes) are for the net returns (not log returns).

Proof. (of Proposition 7.13) Notice that

$$\frac{\mathrm{E}(1+R_p)^{1-\gamma}}{1-\gamma} = \frac{\mathrm{E}\exp[(1-\gamma)r_p]}{1-\gamma}, \text{ where } r_p = \ln(1+R_p).$$

(Clearly, when utility is written in terms of wealth $W_0(1+R_p)$, both sides are multiplied by $W_0^{1-\gamma}$, which does not affect the optimization problem.) Since r_p is normally distributed, the expectation is (recall that if $x \sim N(\mu, \sigma^2)$, then $\mathrm{E}\, e^x = e^{\mu+\sigma^2/2}$)

$$\frac{1}{1-\gamma}\mathrm{E}\exp[(1-\gamma)r_p] = \frac{1}{1-\gamma}\exp[(1-\gamma)\mathrm{E}\, r_p + (1-\gamma)^2 \operatorname{Var}(r_p)/2].$$

Assume that $\gamma > 1$. The function $\ln[z(1-\gamma)]/(1-\gamma)$ is then defined for $z < 0$ and it is increasing in z, see Figure 7.8.b. Let z be the the right hand side of the previous equation and apply the transformation to get

$$\mathrm{E}\, r_p + (1-\gamma)\,\mathrm{Var}(r_p)/2,$$

which is increasing in the expected log return and decreasing in the variance of the log return (since we assumed $1-\gamma < 0$). To express this in terms of the mean and variance of the return instead of the log return we use the following fact: if $\ln y \sim N(\mu, \sigma^2)$, then $\mathrm{E}\, y = \exp(\mu + \sigma^2/2)$ and $\mathrm{Std}\,(y) \,/\, \mathrm{E}\, y = \sqrt{\exp(\sigma^2) - 1}$. Using this fact on the previous expression gives

$$\ln(1 + \mathrm{E}\, R_p) - \gamma \ln[\mathrm{Var}(R_p)/(1 + \mathrm{E}\, R_p)^2 + 1]/2,$$

which is increasing in $\mathrm{E}\, R_p$ and decreasing in $\mathrm{Var}(R_p)$. We therefore get a mean-variance portfolio. ■

7.3 Application of Normal Returns: Value at Risk, ES, Lpm and the Telser Criterion

The mean-variance framework is often criticized for failing to distinguish between downside (considered to be risk) and upside (considered to be potential). This section illustrates that normally distributed returns often lead to minimum variance portfolios even if the portfolio selection model seems to be far from the standard mean-variance utility function.

7.3.1 Value at Risk and the Telser Criterion

If the return is normally distributed, $R \sim N(\mu, \sigma^2)$, then the α value at risk, VaR_α, is

$$\mathrm{VaR}_\alpha = -(\mu + c_{1-\alpha}\sigma), \tag{7.24}$$

where $c_{1-\alpha}$ is the $1-\alpha$ quantile of a N(0,1) distribution, for instance, -1.64 for 5%.

Example 7.14 *(VaR with $R \sim N(\mu, \sigma^2)$) If $\mu = 8\%$ and $\sigma = 16\%$, then $VaR_{95\%} = -(0.08 - 1.64 \times 0.16) \approx 0.18$; we are 95% sure that we will not loose more than 18% of the investment.*

Suppose we abandon MV preferences and instead choose to minimize the Value at Risk—for a given mean return. With normally distributed returns, the value at risk (7.24) is a strictly increasing function of the standard deviation (and the variance). Hence, minimizing the value at risk gives the same solution (portfolio weights) as minimizing the variance. (However, it should be noted that the VaR approach is often used when data is thought to be strongly non-normal.)

Another portfolio choice approach is to use the value at risk as a restriction. For instance, the *Telser criterion* says that we should maximize the expected portfolio return subject to the restriction that the value at risk (at some given probability level) does not exceed a given level.

The restriction could be that the $\text{VaR}_{95\%}$ should be less than 10% of the investment. With a normal distribution, (7.24) says that the portfolio must be such that the mean and standard deviation satisfy

$$\begin{aligned} -(\mu_p - 1.64\sigma_p) < 0.1, \text{ or} \\ \mu_p > -0.1 + 1.64\sigma_p. \end{aligned} \tag{7.25}$$

The portfolio choice problem according to the Telser criterion is then to choose the portfolio weights (v_i) to

$$\max_{v_i} \mu_p \text{ subject to } \mu_p > -0.1 + 1.64\sigma_p \text{ and } \Sigma_{i=1}^n v_i = 1. \tag{7.26}$$

More generally, the Telser criterion is

$$\max_{v_i} \mu_p \text{ subject to } \mu_p > -\text{VaR}_\alpha - c_{1-\alpha}\sigma_p \text{ and } \Sigma_{i=1}^n v_i = 1, \tag{7.27}$$

where $c_{1-\alpha}$ is the $1-\alpha$ quantile of a $N(0,1)$ distribution.

This problem is illustrated in Figure 7.10. Any point above a line satisfies the restriction, and the issue is to pick the one with the highest possible expected return—among those available. In particular, there are no portfolios above the minimum-variance frontier (with or without a riskfree asset). A lower VaR is, of course, a tougher restriction.

If the restriction intersects the minium-variance frontier, the solution is the highest intersection point. This is indeed a point on the minimum-variance frontier, which shows that the Telser criterion applied to normally distributed returns leads us to a minimum-variance portfolio. If the restriction doesn't intersect, then there is no solution to the

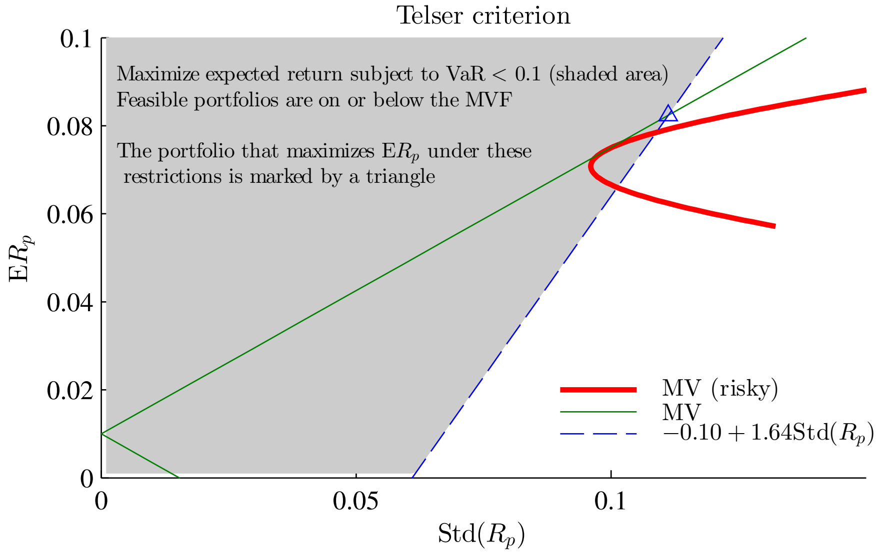


Figure 7.10: Telser criterion and VaR

problem (the restriction is too demanding, the VaR too low).

7.3.2 Expected Shortfall

The expected shortfall is the expected loss when the return actually is below the VaR_α. For normally distributed returns, $R \sim N(\mu, \sigma^2)$, it can be shown that

$$\text{ES}_\alpha = -\mu + \sigma \frac{\phi(c_{1-\alpha})}{1-\alpha}, \tag{7.28}$$

where $\phi()$ is the pdf or a $N(0,1)$ variable.

Example 7.15 *If $\mu = 8\%$ and $\sigma = 16\%$, the 95% expected shortfall is $ES_{95\%} = -0.08 + \sigma\phi(1.64)/0.05 \approx 0.25$.*

Notice that the expected shortfall for a normally distributed return (7.28) is a strictly increasing function of the standard deviation (and the variance). As for the VaR, this means that minimizing expected shortfall at a given mean return therefore gives the same solution (portfolio weights) as minimizing the variance at the same given mean return.

A "Telser criterion" could, for instance, use the restriction $\text{ES}_\alpha < 0.25$

$$\mu_p > -0.25 + \sigma_p \frac{\phi(c_{1-\alpha})}{1-\alpha}, \tag{7.29}$$

which is define an area in a MV figure similar to that in Figure 7.10.

7.3.3 Target Semivariance

Reference: Bawa and Lindenberg (1977) and Nantell and Price (1979)

Using the variance (or standard deviation) as a measure of portfolio risk (as a mean-variance investor does) fails to distinguish between the downside and upside. As an alternative, one could consider using a target semivariance (lower partial 2nd moment) instead. It is defined as

$$\lambda_p(h) = \mathrm{E}[\min(R_p - h, 0)^2], \tag{7.30}$$

where h is a "target level" chosen by the investor. In the subsequent analysis it will be set equal to the riskfree rate.

Suppose investors preferences are such that they like high expected returns and dislike the target semivariance—with a target level equal to the riskfree rate (denoted λ_p to keep the notation brief), that is, if their expected utility can be written as

$$\begin{aligned} \mathrm{E}\,U\left(R_p\right) &= V(\mu_p, \lambda_p), \text{ with} \qquad (7.31) \\ \partial(\mu_p, \lambda_p)/\partial\mu_p > 0 &\text{ and } \partial(\mu_p, \lambda_p)/\partial\lambda_p < 0. \end{aligned}$$

The results in Bawa and Lindenberg (1977) and Nantell and Price (1979) demonstrate several important things. First, there is still a two-fund theorem: all investors hold a combination of a market portfolio and the riskfree asset, so there is a capital market line. See Figure 7.11 for an illustration (based on normally distributed returns, which is not necessary). Second, there is still a beta representation as in CAPM, but where the beta coefficient is different.

Third, in case the returns are normally distributed (or t-distributed), then the optimal portfolios are also on the mean-variance frontier, and all the usual MV results hold. See Figure 7.12 for a numerical illustration.

The basic reason is that $\lambda_p(h)$ is increasing in the standard deviation (for a given mean). This means that minimizing $\lambda_p(h)$ at a given mean return gives exactly the same solution (portfolio weights) as minimizing σ_p (or σ_p^2) at the same given mean return.

As a result, with normally distributed returns, an investor who wants to minimize the target semivariance (at a given mean return) is behaving just like a mean-variance investor.

Remark 7.16 *(Target semivariance calculation for normally distributed variable*) For*

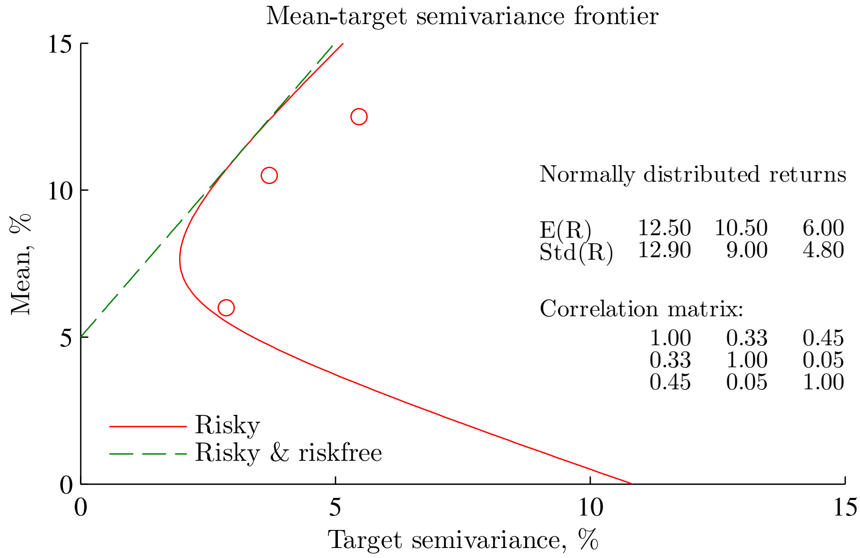


Figure 7.11: Target semivariance and expected returns

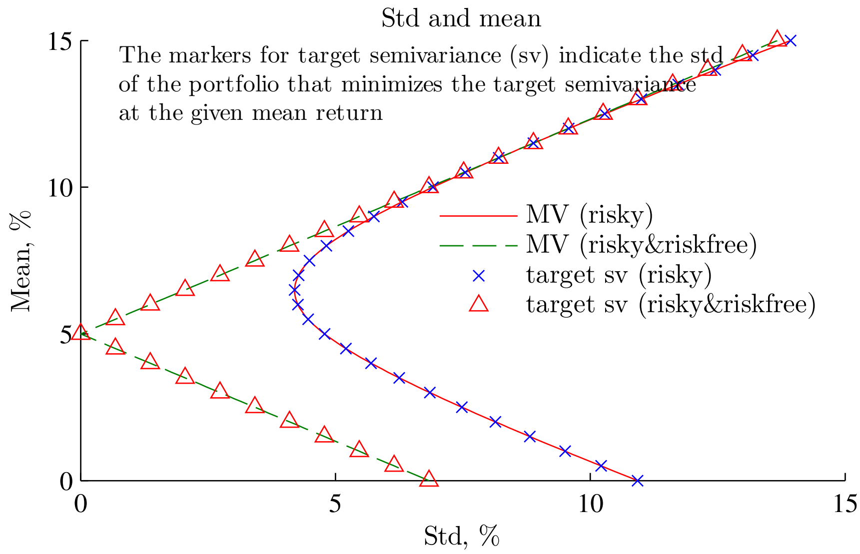


Figure 7.12: Standard deviation and expected returns

an $N(\mu, \sigma^2)$ variable, the target semivariance around the target level h is

$$\lambda_p(h) = \sigma^2 a\phi(a) + \sigma^2(a^2 + 1)\Phi(a), \text{ where } a = (h - \mu)/\sigma,$$

while $\phi()$ and $\Phi()$ are the pdf and cdf of a $N(0,1)$ variable respectively. Notice that $\lambda_p(h) = \sigma^2/2$ for $h = \mu$. It is straightforward to show that

$$\frac{\partial \lambda_p(h)}{\partial \sigma} = 2\sigma \Phi(a),$$

so the target semivariance is a strictly increasing function of the standard deviation.

7.4 Behavioural Finance

Reference: Elton, Gruber, Brown, and Goetzmann (2010) 18; Forbes (2009); Shefrin (2005)

There is relatively little direct evidence on investor's preferences (utility). For obvious reasons, we can't know for sure what people really like. The evidence we do have is from two sources: "laboratory" experiments designed to elicit information about the test subject's preferences for risk, and a lot of indirect information.

7.4.1 Evidence on Utility Theory

The laboratory experiments are typically organized at university campuses (mostly by psychologists and economists) and involve only small compensations—so the test subjects are those students who really need the monetary compensation for taking part or those that are interested in this type of psychological experiments. The results vary quite a bit, but a main theme is that the main assumptions in utility-based portfolio choice might be reasonable, but there are some important systematic deviations from these assumptions.

For instance, investors seem to be unwilling to realize losses, that is, to sell off assets which they have made a loss on (often called the "disposition effect"). They also seem to treat the investment problem much more on an asset-by-asset basis than suggested by mean-variance analysis which pays a lot of attention to the covariance of assets (sometimes called mental accounting). Discounting appears to be non-linear in the sense that discounting is higher when comparing today with dates in the near future than when comparing two dates in the distant future. (Hyperbolic discount factors might be a way to model this, but lead to time-inconsistent behaviour: today we may prefer an asset that pays off in $t+2$ to an asset than pays off in $t+1$, but tomorrow our ranking might be reversed.) Finally, the results seem to move towards tougher play as the experiments are repeated and/or as more competition is introduced—although the experiments seldom converge to ultra tough/egoistic behaviour (as typically assumed by utility theory).

The indirect evidence is broadly in line with the implications of utility-based theory—especially now that the costs for holding well diversified portfolios have decreased (mutual funds). However, there are clearly some systematic deviations from the theoretical implications. For instance, many investors seem to be too little diversified. In particular, many investors hold assets in companies/countries that are very strongly correlated to their labour income (local bias). Moreover, diversification is often done in a naive fashion and depend on the "menu" of choices. For instance, many pension savers seems to diversify by putting the fraction $1/n$ in each of the n funds offered by the firm/bank—irrespective of what kind of funds they are. There are, of course, also large chunks of wealth invested for control reasons rather than for a pure portfolio investment reason (which explains part of the so called "home bias"—the fact that many investors do not diversify internationally).

7.4.2 Evidence on Expectations Formation (Forecasting)

In laboratory experiments (and studies of the properties of forecasts made by analysts), several interesting results emerge on how investors seems to form expectations. First, complex situations are often approached by treating them as a simplified representative problem—even against better knowledge (often called "representativeness")—and stands in contrast to the idea of Bayesian learning where investors update and learn from their mistakes. Second (and fairly similar), difficult problems are often handled as if they were similar to some old/easy problem—and all that is required is a small modification of the logic (called "anchoring"). Third, recent events/data are given much higher weight than they typically warrant (often called "recency bias" or "availability"). Finally, most forecasters seem to be overconfident: they draw too strong conclusions from small data sets ("law of small numbers") and overstate the precision of their own forecasts.

Notice, however, that it is typically difficult to disentangle (distorted) beliefs from non-traditional preferences. For instance, the aversion of selling off bad investments, may equally well be driven by a belief that past losers will recover.

7.4.3 Prospect Theory

The *prospect theory* (developed by Kahneman and Tversky) try to explain several of these things by postulating that the utility function is concave over some reference point (which may shift), but convex below it. This means that gains are treated in a risk averse way, but losses in a risk loving way. For instance, after a loss (so we are below the reference point) an asset looks less risky than after a gain—which might explain why investors hold on to losing investments. Clearly, an alternative explanation is that investors believe in mean-reversion (losing positions will recover, winning positions will fall back). In general, it is hard to make a clear distinction between non-classical preferences and (potentially distorted) beliefs.

Bibliography

Bawa, V. S., and E. B. Lindenberg, 1977, "Capital market equilibrium in a mean-lower partial moment framework," *Journal of Financial Economics*, 5, 189–200.

Cochrane, J. H., 2001, *Asset pricing*, Princeton University Press, Princeton, New Jersey.

Danthine, J.-P., and J. B. Donaldson, 2002, *Intermediate financial theory*, Prentice Hall.

Elton, E. J., M. J. Gruber, S. J. Brown, and W. N. Goetzmann, 2010, *Modern portfolio theory and investment analysis*, John Wiley and Sons, 8th edn.

Forbes, W., 2009, *Behavioural finance*, Wiley.

Huang, C.-F., and R. H. Litzenberger, 1988, *Foundations for financial economics*, Elsevier Science Publishing, New York.

Ingersoll, J. E., 1987, *Theory of financial decision making*, Rowman and Littlefield.

Nantell, T. J., and B. Price, 1979, "An analytical comparison of variance and semivariance capital market theories," *Journal of Financial and Quantitative Analysis*, 14, 221–242.

Shefrin, H., 2005, *A behavioral approach to asset pricing*, Elsevier Academic Press, Burlington, MA.

8 CAPM Extensions

Reference: Elton, Gruber, Brown, and Goetzmann (2010) 14 and 16

8.1 Background Risk

This section discusses the portfolio problem when there is "background risk." For instance, it often makes sense to treat labour income, social security payments and perhaps also real estate as (more or less) background risk. The same applies to the value of a liability stream. A target retirement wealth or planned future house purchase can be thought of as a virtual liability.

The existence of background will typically affect the portfolio choice and therefore also asset prices—at least as long as the background risk is correlated with some assets. The intuition is that the assets will be used to hedge against the background risk.

8.1.1 Portfolio Choice with Background Risk: One Risky Asset

To build a simple example, consider a mean-variance investor who can choose between a riskfree asset (with return R_f) and equity (with return R_1). He also has a background risk—in the form of an endowment (positive or negative) of an asset (with return R_H). This could, for instance, be labour income or a house (positive endowment). For a company, it could perhaps the present value of a liability stream (negative endowment) or the need to buy some commodities to the company's production process next period (also like a negative endowment—from the perspective of the CFO). The investor's portfolio problem is to maximize

$$\mathrm{E}\,U(R_p) = \mathrm{E}\,R_p - \frac{k}{2}\operatorname{Var}(R_p), \text{ where} \tag{8.1}$$

$$R_p = vR_1 + \phi R_H + (1 - v - \phi)R_f \tag{8.2}$$

$$= vR_1^e + \phi R_H^e + R_f. \tag{8.3}$$

Note that ϕ is the portfolio weight of the background risk (which is not a choice variable—rather an "endowment") and $1 - \phi$ is the weight of the financial portfolio (riskfree plus "equity"). Recall that ϕ is negative if the background risk is a liability (so the investor is endowed with a short position in the background risk).

Use the budget constraint in the objective function to get (using the fact that R_f is known)

$$\mathrm{E}\, U(R_p) = v\mu_1^e + \phi\mu_H^e + R_f - \frac{k}{2}\left(v^2\sigma_{11} + \phi^2\sigma_{HH} + 2v\phi\sigma_{1H}\right), \tag{8.4}$$

where σ_{11} and σ_{HH} are the variances of equity and the background risk respectively, and σ_{1H} is their covariance.

The first order condition for the weight on equity, v, is $\partial\, \mathrm{E}\, U(R_p)/\partial v = 0$, that is,

$$\begin{aligned} 0 &= \mu_1^e - k\left(v\sigma_{11} + \phi\sigma_{1H}\right), \text{ so} \\ v &= \frac{\mu_1^e/k - \phi\sigma_{1H}}{\sigma_{11}}. \end{aligned} \tag{8.5}$$

Notice that the second term, $-\phi\sigma_{1H}/\sigma_{11}$ (also called the "hedging term") depends on how important the background is in the portfolio (ϕ) and the "beta" of the background risk from a regression

$$R_H^e = \alpha + \beta R_1^e + \varepsilon, \text{ since } \beta = \sigma_{1H}/\sigma_{11}. \tag{8.6}$$

Essentially, the hedging term is related to how equity can help us create a hedge against the background risk. If the beta is positive, then equity tends to move in the same direction as the background, so a short equity position eliminates a lot of a positive exposure ($\phi > 0$) to the background risk—and vice versa.

It is also interesting that the optimal portfolio weight (8.5) does not depend on the return on the background risk. This might seem somewhat unintuitive. After all, if an investor is rich like a troll (according to Scandinavian legends, trolls are supposed to be rich) then he ought to be able to carry more risk. However, that is not how the mean variances preferences work. Rather, those preferences say something about how much *extra* average returns that are required in order to carry a certain amount of extra volatility. (The answer does not depend on the general level of mean returns since the preferences are linear in both the portfolio mean return and variance.)

The presence of background risk has important consequences for the portfolio weights of the financial subportfolio. This subportfolio has the weights $w = v/(1-\phi)$ on equity and $w_f = (1 - v - \phi)/(1-\phi)$ on the riskfree assets (summing to unity). By using (8.5), these weights are

$$w = \frac{v}{1-\phi} = \frac{\mu_1^e/k - \phi\sigma_{1H}}{(1-\phi)\sigma_{11}} \text{ and} \tag{8.7}$$

$$w_f = 1 - w. \tag{8.8}$$

First, *when the covariance is zero* ($\sigma_{1H} = 0$), then, the equity weight is increasing in the amount of background risk (ϕ), while the opposite holds for the riskfree asset. The intuition is that a zero covariance means that the background risk is quite similar to a bond: having an endowment of a bond-like asset in the overall portfolio means that the financial portfolio should tilted away from actual bonds.

Second, *when the covariance is positive* ($\sigma_{1H} > 0$) and we have a positive exposure to the background risk ($\phi > 0$), then the hedging term (second term) will then tilt the financial portfolio away from equity and towards the safe asset. The intuition is that the overall portfolio now includes a lot of "equity like" assets, so the financial portfolio should be tilted towards bonds. The opposite holds when the exposure to the background risk is negative (a liability, $\phi < 0$) or when the background risk is negatively correlated with equity ($\sigma_{1H} < 0$, assuming a positive exposure, $\phi > 0$).

Example 8.1 *(Portfolio choice with background risk) Suppose* $k = 3, \mu_1^e = 0.08$ *and* $\sigma_{11} = 0.2^2$*, then (8.5) gives*

	v_1	w_1
Case A ($\phi = 0$)	0.67	0.67
Case B ($\phi = 0.5, \sigma_{1H} = 0$)	0.67	1.33
Case C ($\phi = 0.5, \sigma_{1H} = 0.01$)	0.54	1.08

Comparing cases A and B, we see that adding background risk that is uncorrelated with equity tilts the financial portfolio towards equity. Comparing cases B and C, we see that this effect is less pronounced if the background risk is positively correlated with equity.

Example 8.2 *(Portfolio choice with a liability) Continuing Example 8.1, suppose now* *that the background risk is a liability (short position). Then (8.5) gives*

	$\underline{v_1}$	$\underline{w_1}$
Case D ($\phi = -0.5, \sigma_{1H} = 0$)	0.67	0.44
Case E ($\phi = -0.5, \sigma_{1H} = 0.01$)	0.79	0.53

Comparing cases A and D, we see that adding a liability risk risk that is uncorrelated with equity tilts the financial portfolio towards bonds. The reason is that the liability is like a short position in bonds which we cover by buying more actual bonds. Comparing cases D and E, we see that a liability risk that is positively correlated with equity tilts the financial portfolio towards equity. The reason is that the liability is now like a short position in equity which we cover by buying more equity.

Example 8.3 *(Portfolio choice of young and old) Consider the common portfolio advice that* young investors *(with labour income) should invest relatively more in stocks than* old investors *(without labour income). In this case, the background risk is an endowment of "human capital," that is, the present value of future labour income—and current labour income can loosely be interpreted as its return. The analysis in the previous section suggests that a low correlation of stock returns and wages means that the young investor is endowed with a bond-like asset. His financial portfolio will therefore be tilted towards the risky asset—compared to the old investor. (This intuition is strengthened by the fact that labour income is typically a lot less volatile than equity returns.)*

Remark 8.4 *(Optimising over w directly*) Rewrite the portfolio return (8.2) as*

$$\begin{aligned} R_p &= w(1-\phi)R_1 + (1-w)(1-\phi)R_f + \phi R_H \\ &= w(1-\phi)R_1^e + Z_f, \text{ where } Z_f = (1-\phi)R_f + \phi R_H. \end{aligned}$$

Use in the objective function (and notice that Z_f is a risky asset) to get

$$\mathrm{E}\, U(R_p) = w(1-\phi)\mu_1^e + \mu_f - \frac{k}{2}\left[w^2(1-\phi)^2\sigma_{11} + \sigma_{ff} + 2w(1-\phi)\sigma_{1f}\right].$$

The first order condition with respect to w gives

$$\begin{aligned} 0 &= \mu_1^e - k\left[w(1-\phi)\sigma_{11} + \sigma_{1f}\right], \text{ so} \\ w &= \frac{\mu_1^e/k - \sigma_{1f}}{(1-\phi)\sigma_{11}}. \end{aligned}$$

Since $\sigma_{1f} = \text{Cov}(R_1, Z_f) = \phi\sigma_{1H}$*, this is the same as in (8.8).*

8.1.2 Portfolio Choice with Background Risk: Several Risky Assets

With several risky assets the portfolio return is

$$R_p = v'R + (1 - 1'v - \phi)R_f + \phi R_H, \tag{8.9}$$

where v is a vector of portfolio weights, R a vector of returns on the risky assets and 1 is a vector of ones (so $1'v$ is the sum of the elements in the v vector). In this case we get

$$v = \Sigma^{-1}\left(\mu^e/k - \phi S_H\right), \text{ and} \tag{8.10}$$

$$w = v/(1-\phi), \tag{8.11}$$

where Σ is the covariance matrix of all assets and S_H is a vector of covariances of the assets with the background risk.

Proof. (of (8.10)) The investor solves

$$\max_v v'\mu^e + \phi\mu^e_H + R_f - \frac{k}{2}\left(v'\Sigma v + \phi^2\sigma_{HH} + 2\phi v'S_H\right),$$

with first order conditions

$$\mathbf{0} = \mu^e - k\left(\Sigma v + \phi S_H\right), \text{ so}$$
$$v = \Sigma^{-1}\left(\mu^e/k - \phi S_H\right).$$

■

As in the univariate case, the hedging term depends on betas from a regression of R^e_H on the vector of risky assets (R^e)

$$R^e_H = \alpha + \beta' R^e + \varepsilon, \text{ since } \beta = \Sigma^{-1}S_H. \tag{8.12}$$

It can also be noted that the background risk could well be a "portfolio" of different background risks, for instance, labour income plus owning a house (positive) or a planned retirement wealth and future house purchase (negative). The properties of the elements of this portfolio matters only so far as they affect the covariances S_H. The portfolio weights in (8.11) will (as long as $\phi S_H \neq 0$) give a return that is off the mean-variance frontier.

See Figure 8.1 for an illustration.

However, the portfolio is on the *mean-variance frontier of some transformed assets* $Z_i = (1-\phi)R_i + \phi R_H$. In fact, we can rewrite the portfolio return (8.9) as

$$\begin{aligned} R_p &= w'Z + (1-1'w)Z_f, \text{ where} \\ Z_i &= (1-\phi)R_i + \phi R_H. \end{aligned} \tag{8.13}$$

Proof. ((8.13) is the same as (8.9)) Write out (8.13) and simplify

$$\begin{aligned} R_p &= w'\left[(1-\phi)R + \phi R_H\right] + (1-1'w)\left[(1-\phi)R_f + \phi R_H\right] \\ &= (1-\phi)w'R + \phi 1'wR_H + (1-\phi)(1-1'w)R_f + (1-1'w)\phi R_H \\ &= (1-\phi)w'R + (1-\phi)(1-1'w)R_f + \phi R_H. \end{aligned}$$

Let $(1-\phi)w = v$, so the coefficients on R are the same as in (8.9). This definition implies that the coefficient on R_f is $(1-\phi)(1-1'v/(1-\phi)) = (1-\phi-1'v)$ which is also the same as in (8.9). ■

Maximizing the objective function (8.1) subject to this new definition of the portfolio return is a standard mean-variance problem—but in terms of the transformed assets Z_i (which are all risky). Therefore, the optimal portfolio will be on the mean-variance frontier of these transformed assets. See Figure 8.1 for an illustration.

Example 8.5 *(Portfolio choice, two traded assets and background risk) With two risky traded assets and background risk the investor maximizes* $\mathrm{E}\, R_p - \frac{k}{2}\mathrm{Var}(R_p)$*, where* $R_p = v_1 R_1^e + v_2 R_2^e + \phi R_H^e + R_f$*, that is*

$$\max_{v_1,v_2} v_1\mu_1^e + v_2\mu_2^e + \phi\mu_H^e + R_f - \frac{k}{2}\left[v_1^2\sigma_{11} + v_2^2\sigma_{22} + \phi^2\sigma_{HH} + 2v_1v_2\sigma_{12} + 2v_1\phi\sigma_{1H} + 2v_2\phi\sigma_{2H}\right].$$

The first order conditions are

$$\begin{aligned} 0 &= \mu_1^e - k\left[v_1\sigma_{11} + v_2\sigma_{12} + \phi\sigma_{1H}\right] \\ 0 &= \mu_2^e - k\left[v_2\sigma_{22} + v_1\sigma_{12} + \phi\sigma_{2H}\right], \end{aligned}$$

or

$$\begin{bmatrix} \mu_1^e \\ \mu_2^e \end{bmatrix} = k \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + k\phi \begin{bmatrix} \sigma_{1H} \\ \sigma_{2H} \end{bmatrix}.$$

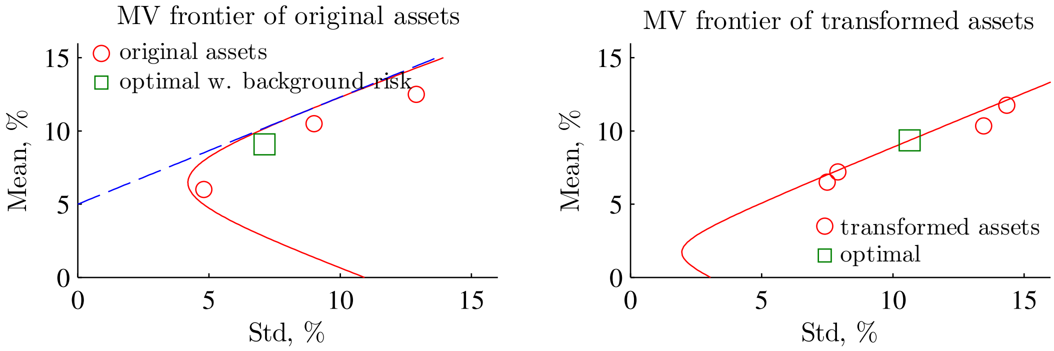


Figure 8.1: Portfolio choice with background risk

The solution is

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix} \left(\begin{bmatrix} \mu_1^e \\ \mu_2^e \end{bmatrix} \frac{1}{k} - \phi \begin{bmatrix} \sigma_{1H} \\ \sigma_{2H} \end{bmatrix} \right).$$

Example 8.6 *(Portfolio choice of a pharmaceutical engineer) In the previous remark, suppose asset 1 is an index of pharmaceutical stocks, and asset 2 is the rest of the equity market. Consider a person working as a pharmaceutical engineer: the covariance of her labour with asset 1 is likely to be high, while the covariance with asset 2 might be fairly small. This person should therefore tilt his financial portfolio away from pharmaceutical stocks: the market portfolio is not the best for everyone.*

8.1.3 Asset Pricing Implications of Background Risk

The beta representation of expected returns is also affected by the existence of background risk. Let R_m denote the market portfolio of the marketable assets (whose weights are proportional to (8.10)). We then have

$$\mu_i^e = \tilde{\beta}_i \mu_m^e, \text{ where } \tilde{\beta}_i = \frac{\sigma_{im} + \phi\,(\sigma_{iH} - \sigma_{im})}{\sigma_{mm} + \phi\,(\sigma_{mH} - \sigma_{mm})}. \tag{8.14}$$

This coincides with the standard case when $\phi = 0$ (no background risk) or when both asset i and the market are uncorrelated with the background risk. This expression suggests one reason for why the traditional beta (against the market portfolio only) could be biased. For instance, if the market is positively correlated with R_H, but asset i is negatively correlated with R_H, then $\tilde{\beta}_i$ is lower than the traditional beta.

Proof. (*of (8.14)) Divide the portfolio weights in (8.10) by $1-\phi$ to get the weights of the (financial) market portfolio, w_m. For any portfolio with portfolio weights w_p we have the covariance with the market

$$\begin{aligned}\sigma_{pm} &= w_p' \Sigma w_m \\ &= w_p' \Sigma \Sigma^{-1} \left(\mu^e / k - S_H \phi\right) / \left(1-\phi\right) \\ &= \mu_p^e / \left[k\left(1-\phi\right)\right] - \sigma_{pH}\phi / \left(1-\phi\right).\end{aligned}$$

Apply this equation to the market return itself to get

$$\sigma_{mm} = \mu_m^e / \left[k\left(1-\phi\right)\right] - \sigma_{mH}\phi / \left(1-\phi\right).$$

Combine these two equations as

$$\frac{\sigma_{pm} + \sigma_{pH}\phi / \left(1-\phi\right)}{\sigma_{mm} + \sigma_{mH}\phi / \left(1-\phi\right)} = \frac{\mu_p^e}{\mu_m^e},$$

which can be rearranged as (8.14). ■

Notice that a standard CAPM regression of

$$R_i^e = \alpha_i + b_i R_m^e + \varepsilon_i, \tag{8.15}$$

would produce (in a very large sample) the traditional beta ($b_i = \beta = \sigma_{im}/\sigma_{mm}$) and a non-zero intercept equal to

$$\alpha_i = (\tilde{\beta}_i - \beta_i)\mu_m^e. \tag{8.16}$$

A rejection of the null that the intercept is zero (a rejection of CAPM) could then be due to the existence of background risk. (There are clearly several other possible reasons.)

Proof. (of (8.16)) Take expectations of (8.15) to get $\mu_i^e = \alpha_i + \beta_i \mu_m^e$. From (8.14) we then have $\tilde{\beta}_i \mu_m^e = \alpha_i + \beta_i \mu_m^e$ which gives (8.16). ■

Example 8.7 *(Different betas) Suppose $\sigma_{im} = 0.8, \sigma_{mm} = 1, \sigma_{iH} = -0.5$, and $\sigma_{mH} =$* 0.5

$$\tilde{\beta}_i = \begin{cases} \frac{0.8}{1} = 0.8 & \text{if } \phi = 0 \\ \frac{0.8+0.3(-0.5-1)}{1+0.3(0.5-1)} = 0.41 & \text{if } \phi = 0.3. \end{cases}$$

There is also another way to express the expected excess return of asset i—as a *multi-factor model* (or multi-beta model).

$$\mu_i^e = \beta_{im}\mu_m^e + \beta_{iH}\mu_H^e. \tag{8.17}$$

In this case, the expected excess return on asset i depends on how it is related to both the (financial) market and the background risk. The key implication of (8.17) is that there are two risk factors that influence the required risk premium of asset i: both the market and the background risk matter. The investor's portfolio choice will typically depend on the background risk, which in turn will affect asset prices (and returns).

It may seem as if we now have a paradox: both the "adjusted" single-beta representation (8.14) and the multiple-beta representation (8.17) are supposedly true. Can that really be the case—and how should we then test the model? Well, both expressions are true—but there is a key difference: the betas in (8.17) could be estimated by a multiple regression, whereas $\tilde{\beta}_i$ in (8.14) could not.

Proof. (*of (8.17)) The first equation of the Proof of (8.14) can be written

$$\begin{aligned}
\mu_p^e/k &= (1-\phi)\,\sigma_{pm} + \phi\sigma_{pH} && (*)\\
&= \begin{bmatrix} 1-\phi & \phi \end{bmatrix} \begin{bmatrix} \sigma_{pm} \\ \sigma_{pH} \end{bmatrix} \\
&= \begin{bmatrix} 1-\phi & \phi \end{bmatrix} \begin{bmatrix} \sigma_{mm} & \sigma_{mH} \\ \sigma_{mH} & \sigma_{HH} \end{bmatrix} \begin{bmatrix} \sigma_{mm} & \sigma_{mH} \\ \sigma_{mH} & \sigma_{HH} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{pm} \\ \sigma_{pH} \end{bmatrix} \\
&= \begin{bmatrix} 1-\phi & \phi \end{bmatrix} \begin{bmatrix} \sigma_{mm} & \sigma_{mH} \\ \sigma_{mH} & \sigma_{HH} \end{bmatrix} \begin{bmatrix} \beta_{pm} \\ \beta_{pH} \end{bmatrix} \\
&= \begin{bmatrix} (1-\phi)\,\sigma_{mm} + \phi\sigma_{mH} & (1-\phi)\,\sigma_{mH} + \phi\sigma_{HH} \end{bmatrix} \begin{bmatrix} \beta_{pm} \\ \beta_{pH} \end{bmatrix}. && (**)
\end{aligned}$$

The third line just multiplies and divides by the covariance matrix. The fourth line follows from the usual definition of regression coefficients, $\beta = \text{Var}(x)^{-1}\,\text{Cov}(x, y)$.

Apply the first equation (*) on the market return and an asset with the same return as the R_H (this is a short cut, it would be more precise to use a "factor mimicking" portfolio—it is just a bit more complicated). We then get

$$\begin{aligned}\mu_m^e/k &= (1-\phi)\,\sigma_{mm} + \phi\sigma_{mH} \text{ and}\\ \mu_H^e/k &= (1-\phi)\,\sigma_{mH} + \phi\sigma_{HH}.\end{aligned}$$

Use these to substitute for the row vector in (**) to get

$$\mu_p^e/k = \begin{bmatrix}\mu_m^e/k & \mu_H^e/k\end{bmatrix}\begin{bmatrix}\beta_{pm}\\ \beta_{pH}\end{bmatrix},$$

which is the same as (8.17). ■

8.2 Heterogenous Investors

This section gives a simple example of a model where the investors have different beliefs.

Recall the simple MV problem where investor i solves

$$\max_\alpha \mathrm{E}_i\, R_p - \mathrm{Var}_i(R_p)k_i/2, \text{ subject to} \tag{8.18}$$

$$R_p = \alpha R_m^e + R_f. \tag{8.19}$$

In these expressions, the expectations, variance, and the risk aversion parameter all carry the subscript i to indicate that they may differ between investors. The solution is that the weight on the risky asset is

$$\alpha_i = \frac{1}{k_i}\frac{\mathrm{E}_i\, R_m^e}{\mathrm{Var}_i(R_m^e)}, \tag{8.20}$$

where $\mathrm{E}_i\, R_m^e$ is the investor's expectation of the excess return of the risky asset and $\mathrm{Var}_i(R_m^e)$ the investor's perceived variance.

If all investors have the same initial wealth, then the average (across investors) α_i must be unity—since the riskfree asset is in zero net supply. Suppose there are N investors, then the average of (8.20) is

$$1 = \frac{1}{N}\sum\nolimits_{i=1}^{N}\frac{1}{k_i}\frac{\mathrm{E}_i\, R_m^e}{\mathrm{Var}_i(R_m^e)}. \tag{8.21}$$

This is an equilibrium condition that must hold. We consider a few illustrative special cases.

First, suppose all investors have the same expectations and assessments of the variance, but different risk aversions, k_i. Then, (8.21) can be rearranged as

$$\mathrm{E}\, R_m^e = \tilde{k}\, \mathrm{Var}(R_m^e), \text{ where } \tilde{k} = \frac{1}{\frac{1}{N}\sum_{i=1}^{N} \frac{1}{k_i}}. \tag{8.22}$$

This shows that the risk premium on the market is increasing in the volatility and $\tilde{k}$. The latter is not the average risk aversion, but closely related to it. For instance, if all k_i is scaled up by a factor b so is $\tilde{k}$ (and therefore the risk premium).

Example 8.8 *("Average" risk aversion) If half of the investors have $k = 2$ and the other half has $k = 3$, then $\tilde{k} = 2.4$.*

Second, suppose now that only the expected excess return is the same for all investors. Then, (8.21) can be rearranged as

$$\mathrm{E}\, R_m^e = \frac{1}{\frac{1}{N}\sum_{i=1}^{N} \frac{1}{k_i \mathrm{Var}_i(R_m^e)}}. \tag{8.23}$$

The market risk premium is now increasing in a complicated expression that is closely related to a weighted average of the perceived market variances—where the weights are increasing in the risk aversion. If all variances or risk aversions are scaled up by a factor b so is the risk premium.

Third, suppose only the expected excess returns differ. Then, (8.21) can be rearranged as

$$\frac{1}{N}\sum_{i=1}^{N} \mathrm{E}_i\, R_m^e = k\, \mathrm{Var}(R_m^e). \tag{8.24}$$

Clearly, the average expected excess return is increasing in the risk aversion and variance. To interpret this a bit more, let the return be the capital gain (assuming no dividend in the next period), $R_m = P_{t+1}/P_t$ where the current period is t

$$\frac{1}{N}\sum_{i=1}^{N} \mathrm{E}_i \left(\frac{P_{t+1}}{P_t} - R_f \right) = k\, \mathrm{Var}(R_m^e) \text{ or} \tag{8.25}$$

$$P_t = \frac{1}{k\, \mathrm{Var}(R_m^e) + R_f} \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}_i\, (P_{t+1}). \tag{8.26}$$

This shows that today's market price, P_t, is simply the average expected future price—scaled down by the risk aversion, volatility and the riskfree rate (to create a capital gain to compensate for the risk and the alternative return).

These special cases suggest that, although the general expression (8.21) is complicated, we are unlikely to commit serious errors by sticking to the formulation

$$\mathrm{E}\, R_m^e = k \,\mathrm{Var}(R_m^e), \tag{8.27}$$

as long as we interpret the components as (close to) averages across investors.

8.3 CAPM without a Riskfree Rate*

This section states the main result for CAPM when there is no riskfree asset. It uses two basic ingredients.

First, suppose investors behave as if they had mean-variance preferences, so they choose portfolios on the mean-variance frontier (of risky assets only). Different investors may have different portfolios, but they are all on the mean-variance frontier. The market portfolio is a weighted average of these individual portfolios, and therefore itself on the mean-variance frontier. (Linear combinations of efficient portfolios are also efficient.)

Second, consider the market portfolio. We know that we can find some other efficient portfolio (denote it R_z) that has a zero covariance (beta) with the market portfolio, $\mathrm{Cov}(R_m, R_z) = 0$. (Such a portfolio can actually be found for any efficient portfolio, not just the market portfolio.) Let v_m be the portfolio weights of the market portfolio, and Σ the variance-covariance matrix of all assets. Then, the portfolio weights v_z that generate R_z must satisfy $v_m' \Sigma v_z = 0$ and $v_z' \mathbf{1} = 1$ (sum to unity). The intuition for how the portfolio weights of the R_z assets is that some of the weights have the same sign as in the market portfolio (contributing to a positive covariance) and some other have the opposite sign compared to the market portfolio (contributing to a negative covariance). Together, this gives a zero covariance.

See Figure 8.2 for an illustration.

The main result is then the "zero-beta" CAPM

$$\mathrm{E}(R_i - R_z) = \beta_i \,\mathrm{E}(R_m - R_z). \tag{8.28}$$

Proof. (*of (8.28)) An investor (with initial wealth equal to unity) chooses the portfo-

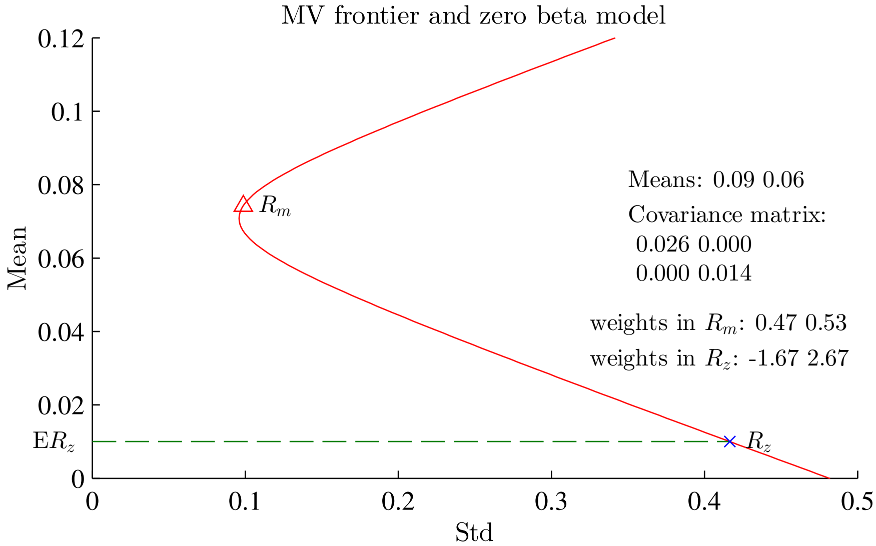


Figure 8.2: Zero-beta model

lio weights (v_i) to maximize

$$\begin{aligned} \mathrm{E}\,U(R_p) &= \mathrm{E}\,R_p - \frac{k}{2}\operatorname{Var}(R_p), \text{ where} \\ R_p &= v_1 R_1 + v_2 R_2 \text{ and } v_1 + v_2 = 1, \end{aligned}$$

where we assume two risky assets. Combining gives the Lagrangian

$$L = v_1\mu_1 + v_2\mu_2 - \frac{k}{2}\left(v_1^2\sigma_{11} + v_2^2\sigma_{22} + 2v_1v_2\sigma_{12}\right) + \lambda(1 - v_1 - v_2).$$

The first order conditions (for v_1 and v_2) are that the partial derivatives equal zero

$$\begin{aligned} 0 &= \partial L/\partial v_1 = \mu_1 - k\,(v_1\sigma_{11} + v_2\sigma_{12}) - \lambda \\ 0 &= \partial L/\partial v_2 = \mu_2 - k\,(v_2\sigma_{22} + v_1\sigma_{12}) - \lambda \\ 0 &= \partial L/\partial \lambda = 1 - v_1 - v_2 \end{aligned}$$

Notice that

$$\sigma_{1m} = \operatorname{Cov}(R_1, \underbrace{v_1R_1 + v_2R_2}_{R_m}) = v_1\sigma_{11} + v_2\sigma_{12},$$

and similarly for σ_{2m}. We can then rewrite the first order conditions as

$$
\begin{aligned}
0 &= \mu_1 - k\sigma_{1m} - \lambda && \text{(a)} \\
0 &= \mu_2 - k\sigma_{2m} - \lambda \\
0 &= 1 - v_1 - v_2
\end{aligned}
$$

Take a weighted average of the first two equations with the weights v_1 and v_2 respectively

$$
\begin{aligned}
v_1\mu_1 + v_2\mu_2 - \lambda &= k\,(v_1\sigma_{1m} + v_2\sigma_{2m}) \\
\mu_m - \lambda &= k\sigma_{mm}, && \text{(b)}
\end{aligned}
$$

which follows from the fact that

$$
\begin{aligned}
v_1\sigma_{1m} + v_2\sigma_{2m} &= v_1 \operatorname{Cov}(R_1, v_1R_1 + v_2R_2) + v_2 \operatorname{Cov}(R_2, v_1R_1 + v_2R_2) \\
&= \operatorname{Cov}(v_1R_1 + v_2R_2, v_1R_1 + v_2R_2) \\
&= \operatorname{Var}(R_m).
\end{aligned}
$$

Divide (a) by (b)

$$
\begin{aligned}
\frac{\mu_1 - \lambda}{\mu_m - \lambda} &= \frac{k\sigma_{1m}}{k\sigma_{mm}} \text{ or} \\
\mu_1 - \lambda &= \beta_1(\mu_m - \lambda)
\end{aligned}
$$

Applying this equation on a return R_z with a zero beta (against the market) gives.

$$\mu_z - \lambda = 0(\mu_m - \lambda), \text{ so we notice that } \lambda = \mu_z.$$

Combining the last two equations gives (8.28). ■

8.4 Multi-Factor Models and APT

8.4.1 Multi-Factor Models

A multi-factor model extends the market model by allowing more factors to explain the return on an asset. In terms of excess returns it could be

$$\begin{aligned} R_i^e &= \beta_{im} R_m^e + \beta_{iF} R_F^e + \varepsilon_i, \text{ where} \\ \mathrm{E}\,\varepsilon_i &= 0, \mathrm{Cov}(R_m^e, \varepsilon_i) = 0, \mathrm{Cov}(R_F^e, \varepsilon_i) = 0. \end{aligned} \tag{8.29}$$

The pricing implication is a multi-beta model

$$\mu_i^e = \beta_{im} \mu_m^e + \beta_{iF} \mu_F^e. \tag{8.30}$$

Remark 8.9 *(When factors are not excess returns) This formulation assumes that the factor can be expressed as an excess return—but that is not necessary. For instance, it could be that the second factor is a macro variable like inflation surprises. Then there are two possible ways to proceed. First, find that portfolio which mimics the movements in the inflation surprises best and use the excess return of that (factor mimicking) portfolio in (8.29) and (8.30). Second, we could instead reformulate the model by adding an intercept in (8.30) and let R_F^e denote whatever the factor is (not necessarily an excess return) and then estimate the factor risk premium, corresponding to μ_F^e in (8.30), by using a cross-section of different assets ($i = 1, 2, \ldots$).*

We have already seen one theoretical multi-factor model: the "CAPM with background risk" in (8.17). The consumption-based model (discussed later on) gives another example. There are also several empirically motivated multi-factor models, that is, empirical models that have been found to work well (even if the theoretical foundation might be a bit weak).

Fama and French (1993) estimate a multi-factor model and show that it performs much better than CAPM. The three factors are: the market return, the return on a portfolio of small stocks minus the return on a portfolio of big stocks, and the return on a portfolio with a high ratio of book value to market value minus the return on a portfolio with a low ratio. He and Ng (1994) try to relate these factors to macroeconomic series.

The multi-factor model by MSCIBarra is widely used in the financial industry. It uses a set of firm characteristics (rather than macro variables) as factors, for instance, size, volatility, price momentum, and industry/country (see Stefek (2002)). This model is often used to value firms without a price history (for instance, before an IPO) or to find mispriced assets.

The APT model (see below) is another motivation for why a multi-factor model may make sense. Finally, consumption-based models typically also suggest multi-factor models (in terms of macro variables).

8.4.2 The Arbitrage Pricing Model

The *first assumption* of the Arbitrage Pricing Theory (APT) is that the return of asset i can be described as

$$\begin{aligned} R_{it} &= a_i + \beta_i f_t + \varepsilon_{i,t}, \text{ where} \qquad (8.31) \\ \mathrm{E}\,\varepsilon_{it} &= 0, \mathrm{Cov}(\varepsilon_{it}, f_t) = \mathrm{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0. \end{aligned}$$

In this particular formulation there is only one factor, f_t, but the APT allows for more factors. Notice that (8.31) assumes that any correlation of two assets (i and j) is due to movements in f_t—the residuals are assumed to be uncorrelated. This is clearly an index model (here a single index).

The *second assumption* of APT is that there are financial markets are very well developed—so well developed that it is possible to form portfolios that "insure" against almost all possible outcomes. To be precise, the assumption is that it is possible to form a zero cost portfolio (buy some, sell some) that has a zero sensitivity to the factor and also (almost) no idiosyncratic risk. In essence, this assumes that we can form a (non-trivial) zero-cost portfolio of the risky assets that is riskfree. In formal terms, the assumption is that there is a non-trivial portfolio (with the value v_j of the position in asset j) such that $\Sigma_{i=1}^{N} v_i = \Sigma_{i=1}^{N} v_i \beta_i = 0$ and $\Sigma_{i=1}^{N} v_i^2 \operatorname{Var}(\varepsilon_{i,t}) \approx 0$. The requirement that the portfolio is non-trivial means that at least some $v_j \neq 0$.

Together, these assumptions imply that (the proof isn't all that simple) for well diversified portfolios we have

$$\mathrm{E}\,R_{it} = R_f + \beta_i \lambda, \qquad (8.32)$$

where λ is (typically) an unknown constant. The important feature is that there is a linear relation between the risk premium (expected excess return) of an asset and its beta. This expression generalizes to the multi-factor case.

Example 8.10 *(APT with three assets) Suppose there are three well-diversified portfolios (that is, with no residual) with the following factor models*

$$\begin{aligned} R_{1,t} &= 0.01 + 1f_t \\ R_{2,t} &= 0.01 + 0.25f_t, \text{ and} \\ R_{3,t} &= 0.01 + 2f_t. \end{aligned}$$

APT then holds if there is a portfolio with v_i invested in asset i, so that the cost of the portfolio is zero (which implies that the weights must be of the form v_1, v_2, and $-v_1 - v_2$ respectively) such that the portfolio has zero sensitivity to f_t, that is

$$\begin{aligned} 0 &= v_1 \times 1 + v_2 \times 0.25 + (-v_1 - v_2) \times 2 \\ &= v_1 \times (1 - 2) + v_2 \times (0.25 - 2) \\ &= -v_1 - v_2 \times 1.75. \end{aligned}$$

There is clearly an infinite number of such weights but they all obey the relation $v_1 = -v_2 \times 1.75$. Notice the requirement that there is no idiosyncratic volatility is (here) satisfied by assuming that none of the three portfolios have any idiosyncratic noise.

Example 8.11 *(APT with two assets) Example 8.10 would not work if we only had the first two assets. To see that, the portfolio would then have to be of the form $(v_1, -v_1)$ and it is clear that $v_1 \times 1 - v_1 \times 0.25 = v_1(1 - 0.25) \neq 0$ for any non-trivial portfolio (that is, with $v_1 \neq 0$).*

One of the main drawbacks with APT is that it is silent about both the number of factors and their definition. In many empirical implications, the factors—or the factor mimicking portfolios—are found by some kind of statistical method. The idea is (typically) to find that combination of some given assets that explain most of the covariance of the same assets. Then, we find the next combination of the same assets that is uncorrelated with the first combination but also explain as much as possible of the (remaining) covariance—and so forth. A few such factors are often enough to account for most of the covariance. Still, the factors have no particular economic interpretation, and it is not possible to guess what the betas ought to be. To do that, we have to get back to the multi-factor model. For instance. CAPM gives the same type of implication as (8.32)—except that CAPM identifies λ as the expected excess return on the market.

8.5 Joint Portfolio and Savings Choice

8.5.1 Two-Period Problem

The basic consumption-based multi-period problem postulates that the investor derives utility from consumption in every period and that the utility in one period is additively separable from the utility in other periods. For instance, if the investor plans for 2 periods (labelled 1 and 2), then he/she chooses the amount invested in different assets to maximize expected utility

$$\max u(C_1) + \delta \operatorname{E}_1 u(C_2), \text{ subject to} \tag{8.33}$$

$$C_1 + I_1 = W_1 \tag{8.34}$$

$$C_2 + I_2 = \left(1 + R_p\right) I_1 + y_2, \text{ where } R_p = v_1 R_1^e + v_2 R_2^e + R_f. \tag{8.35}$$

In equation (8.33) C_t is consumption in period t. The current period (when the portfolio is chosen) is period 1—so all expectations are made on the basis of the information available in period 1. The constant δ is the time discounting, with $0 < \delta < 1$ indicating impatience. (In equilibrium without risk, we will get a positive real interest rate if investors are impatient.)

Equation (8.34) is the budget constraint for period 1: an initial wealth at the beginning of period 1, W_1, is split between consumption, C_1, and investment, I_1. Equation (8.35) is the budget constraint for period 2: consumption plus investment must equal the wealth at the beginning of period 2 plus (exogenous) income, y_2. It is clear that $I_2 = 0$ since investing in period 2 is the same as wasting resources. The wealth at the beginning of period 2 equals the investment in period 1, I_1, times the gross portfolio return—which in turn depends on the portfolio weights chosen in period 1 (v_1 and v_2) as well as on the returns on the assets (from holding them from period 1 to period 2).

Use the budget constraints and $I_2 = 0$ to substitute for C_1 and C_2 in (8.33) to get

$$\max u\left(W_1 - I_1\right) + \delta \operatorname{E}_1 u\left[\left(1 + v_1 R_1^e + v_2 R_2^e + R_f\right) I_1 + y_2\right]. \tag{8.36}$$

The decision variables in period 1 are how much to invest, I_1, (which implicitly defines how much we consume in period 1), and the portfolio weights v_1 and v_2.

The first order condition for I_1 is that the derivative of (8.36) wrt I_1 is zero

$$-u'(C_1) + \delta \, \mathrm{E}_1 \left[u'(C_2) \left(1 + R_p\right) \right] = 0, \tag{8.37}$$

where $u'(C_t)$ is the marginal utility in period t. (In this expression, the consumption levels and the portfolio return are substituted back—in order to facilitate the interpretation.) This says that consumption should be planned so that the marginal loss of utility from investing (decreasing C_1) equals the discounted expected marginal gain of utility from increasing C_2 by the gross return of the money saved.

We can also rewrite (8.37) as

$$\mathrm{E}_1 \left[\frac{\delta u'(C_2)}{u'(C_1)} \left(1 + R_p\right) \right] = 1. \tag{8.38}$$

Since marginal utility is decreasing in consumption (convex utility function), this ratio is increasing in C_1/C_2. Therefore a high portfolio return will be associated with a low C_1/C_2 ratio. As a special case, suppose the investor holds only riskfree assets ($v_1 = v_2 = 0$). The portfolio return is then R_f and is non-random so we can write

$$\mathrm{E}_1 \frac{\delta u'(C_2)}{u'(C_1)} = \frac{1}{1 + R_f} \text{ (if } v_i = 0\text{)}. \tag{8.39}$$

With a high riskfree rate, C_1/C_2 will be low, since it is wortwhile to save.

The first order conditions for v_1 and v_2 are

$$\mathrm{E}_1 \, u'(C_2) R_1^e = 0 \text{ and} \tag{8.40}$$

$$\mathrm{E}_1 \, u'(C_2) R_2^e = 0, \tag{8.41}$$

which say that both excess returns should be orthogonal to marginal utility. To solve for the decision variables (I_1, v_1, v_2) we should use the budget restrictions (8.34) and (8.35) to substitute for C_1 and C_2 in (8.37), (8.40) and (8.41)—and then solve the three equations for the three unknowns. There are typically no explicit solutions, so numerical solutions are the best we can hope for.

The first order conditions still contain some useful information. In particular, recall

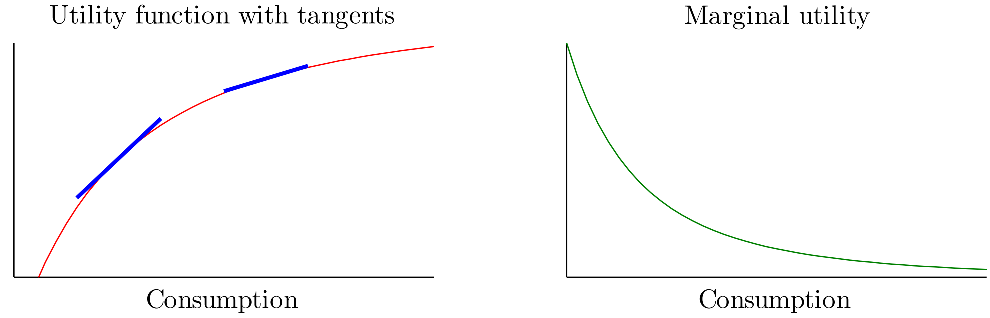


Figure 8.3: Utility function

that, by definition, $\text{Cov}(x, y) = \text{E}\, xy - \text{E}\, x \times \text{E}\, y$, so (8.40) can be written

$$\text{Cov}\left[u'(C_2), R_1^e\right] + \text{E}\, u'(C_2) \times \text{E}\, R_1^e = 0 \text{ or}$$
$$\text{E}\, R_1^e = \frac{\text{Cov}\left[-u'(C_2), R_1^e\right]}{\text{E}\, u'(C_2)}. \tag{8.42}$$

This says that asset 1 will have a high risk premium (expected excess return) if it is negatively correlated with marginal utility, that is, if it tends to have a high return when the need is low. Since marginal utility is decreasing in consumption (concave utility function), this is the same as saying that assets that tend to have high returns when consumption is high (and vice versa) will be considered risky assets—and therefore carry large risk premia. The reason why risky assets have high risk premia is, of course, that otherwise no one would like to buy those assets. (Effectively, high risk means a low price of the asset, so a high dividend yield will contribute to a high average return.) In short, procyclical assets are risky—and will have high expected returns.

Although these results were derived from a two-period problem, it can be shown that a problem with more periods gives the same first-order conditions. In this case, the objective function is

$$u(C_1) + \delta\, \text{E}_1\, u(C_2) + \delta^2\, \text{E}_1\, u(C_3) + \ldots \delta^{T-1}\, \text{E}_1\, u(C_T). \tag{8.43}$$

8.5.2 From a Consumption-Based Model to CAPM

Suppose marginal utility is an affine function of the market excess return

$$u'(C_2) = a - bR_m^e, \text{ with } b > 0. \tag{8.44}$$

This would, for instance, be the case in a Lucas model where consumption equals the market return and the utility function is quadratic–but it could be true in other cases as well. We can then write (8.42) as

$$\mathrm{E}\, R_1^e = b\frac{\mathrm{Cov}\left(R_m^e, R_1^e\right)}{\mathrm{E}\left(a - bR_m^e\right)}. \tag{8.45}$$

We can, of course, apply this expression to the market excess return (instead of asset 1) to get

$$\mathrm{E}\, R_m^e = b\frac{\mathrm{Var}\left(R_m^e\right)}{\mathrm{E}\left(a - bR_m^e\right)}. \tag{8.46}$$

Use (8.46) in (8.45) to substitute $\mathrm{E}\, R_m^e / \mathrm{Var}\left(R_m^e\right)$ for $b/\,\mathrm{E}\left(a - bR_m^e\right)$

$$\mathrm{E}\, R_1^e = \frac{\mathrm{Cov}\left(R_m^e, R_1^e\right)}{\mathrm{Var}\left(R_m^e\right)}\mathrm{E}\, R_m^e, \tag{8.47}$$

which is the beta representation of CAPM.

8.5.3 From a Consumption-Based Model to a Multi-Factor Model

The consumption-based model may not look like a factor model, but it could easily be written as one. The idea is to assume that marginal utility is a linear function of some key macroeconomic variables, for instance, output and interest rates

$$-u'(C_2) = ay + bi. \tag{8.48}$$

Such a formulation makes a lot of sense in most macro models—at least as an approximation. It is then possible to write (8.42) as

$$\mathrm{E}\, R_1^e = \frac{a\,\mathrm{Cov}\left(y, R_1^e\right) + b\,\mathrm{Cov}\left(i, R_1^e\right)}{-\,\mathrm{E}\,(ay + bi)}. \tag{8.49}$$

This, in turn, is easily put in the form of (8.30), where the risk premium on asset 1 depends on the betas against GDP and the interest rate. (See the proof of (8.17) for an idea of how to construct this beta representation.)

8.6 Testing Multi-Factors Models

Provided all factors are excess returns, we can test a multi-factor model by testing if $\alpha = 0$ in the regression

$$R^e_{it} = \alpha + b_{io} R^e_{ot} + b_{ip} R^e_{pt} + ... + \varepsilon_{it}. \tag{8.50}$$

The t-test of the null hypothesis that $\alpha_i = 0$ uses the fact that, under fairly mild conditions, the t-statistic has an asymptotically normal distribution, that is

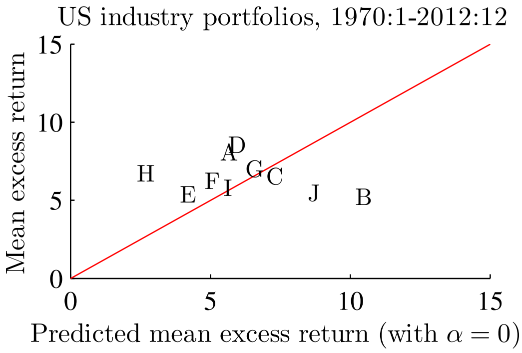
$$\frac{\hat{\alpha}_i}{\text{Std}(\hat{\alpha}_i)} \overset{d}{\to} N(0,1) \text{ under } H_0 : \alpha_i = 0. \tag{8.51}$$

Fama and French (1993) try a multi-factor model. They find that a three-factor model fits the 25 stock portfolios fairly well (two more factors are needed to also fit the seven bond portfolios that they use). The three factors are: the market return, the return on a portfolio of small stocks minus the return on a portfolio of big stocks (SMB), and the return on a portfolio with high BE/ME minus the return on portfolio with low BE/ME (HML). This three-factor model is rejected at traditional significance levels, but it can still capture a fair amount of the variation of expected returns.

Remark 8.12 *(Returns on long-short portfolios*) Suppose you invest x USD into asset i, but finance that by short-selling asset j. (You sell enough of asset j to raise x USD.) The net investment is then zero, so there is no point in trying to calculate an overall return like "value today/investment yesterday - 1." Instead, the convention is to calculate an excess return of your portfolio as $R_i - R_j$ (or equivalently, $R^e_i - R^e_j$). This excess return essentially says: if your exposure (how much you invested) is x, then you have earned $x(R_i - R_j)$. To make this excess return comparable with other returns, you add the riskfree rate: $R_i - R_j + R_f$, implicitly assuming that your portfolio consists includes a riskfree investment of the same size as your long-short exposure (x).*

Chen, Roll, and Ross (1986) use a number of macro variables as factors—along with traditional market indices. They find that industrial production and inflation surprises are priced factors, while the market index might not be.

Figure 8.4 shows some results for the Fama-French model on US industry portfolios and Figures 8.5–8.7 on the 25 Fama-French portfolios.



	alpha	pval	StdErr
all	NaN	0.00	NaN
A (NoDur)	2.70	0.04	8.47
B (Durbl)	-4.96	0.01	12.36
C (Manuf)	-0.46	0.62	6.08
D (Enrgy)	2.90	0.19	14.30
E (HiTec)	1.46	0.34	9.96
F (Telcm)	1.43	0.41	10.86
G (Shops)	0.76	0.60	9.40
H (Hlth)	4.35	0.01	10.66
I (Utils)	0.33	0.85	10.45
J (Other)	-3.00	0.00	5.86

Fama-French model
Factors: US market, SMB (size), and HML (book-to-market)
alpha and StdErr are in annualized %

Figure 8.4: Fama-French regressions on US industry indices

Bibliography

Chen, N.-F., R. Roll, and S. A. Ross, 1986, "Economic forces and the stock market," *Journal of Business*, 59, 383–403.

Elton, E. J., M. J. Gruber, S. J. Brown, and W. N. Goetzmann, 2010, *Modern portfolio theory and investment analysis*, John Wiley and Sons, 8th edn.

Fama, E. F., and K. R. French, 1993, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3–56.

He, J., and L. Ng, 1994, "Economic forces and the stock market," *Journal of Business*, 4, 599–609.

Stefek, D., 2002, "The Barra integrated model," *Barra Research Insight*.

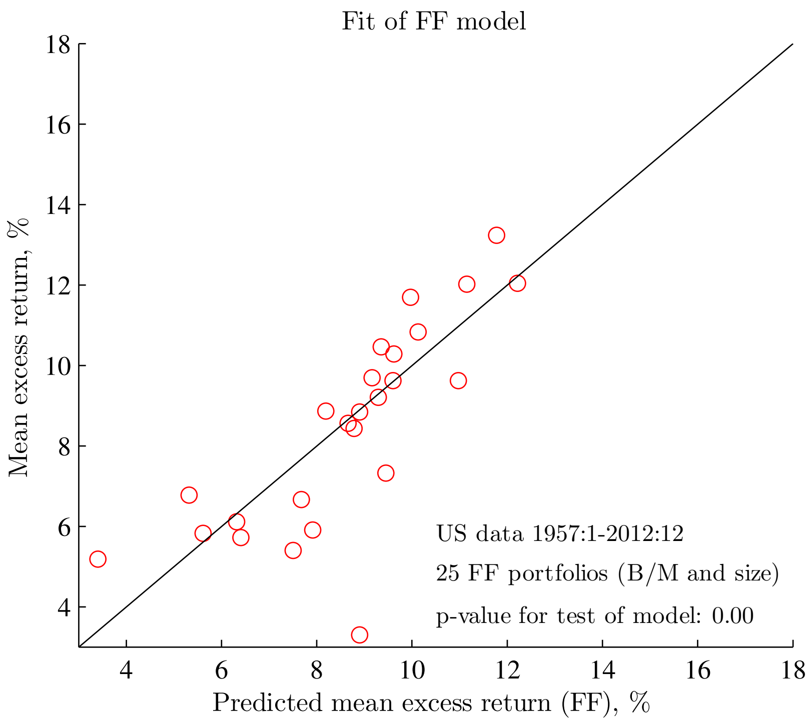


Figure 8.5: FF, FF portfolios

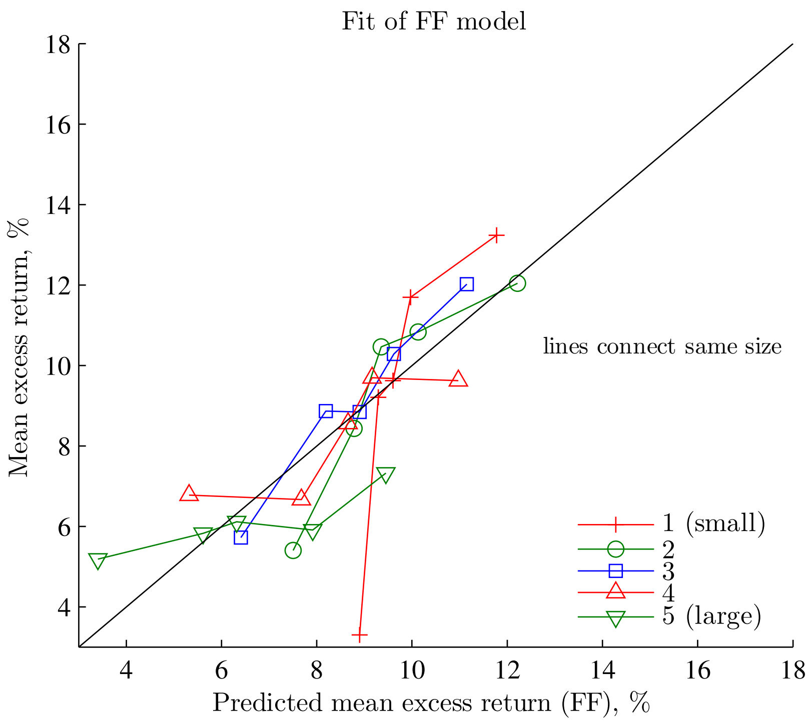


Figure 8.6: FF, FF portfolios

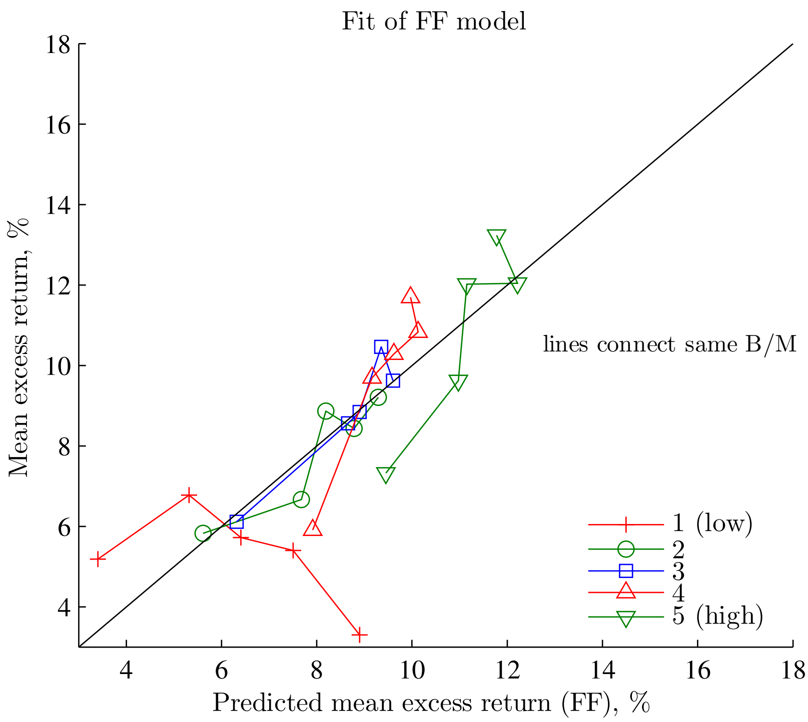


Figure 8.7: FF, FF portfolios

9 Investment for the Long Run

Reference: Campbell and Viceira (2002), Elton, Gruber, Brown, and Goetzmann (2010) 12

9.1 Time Diversification: Approximate Case

This section discusses the notion of "time diversification," which essentially amounts to claiming that equity is safer for long run investors than for short run investors. The argument comes in two flavours: that Sharpe ratios are increasing with the investment horizon, and that the probability that equity returns will outperform bond returns increases with the horizon. This is illustrated in Figure 9.2. The results presented in this section are approximate, since we work with simple returns (and disregard compounding). This has clear disadvantages, but also the advantage of delivering simple results.

9.1.1 Increasing Sharpe Ratios

With iid returns, the expected return and variance both grow linearly with the horizon, so Sharpe ratios (expected excess return divided by the standard deviation) increase with

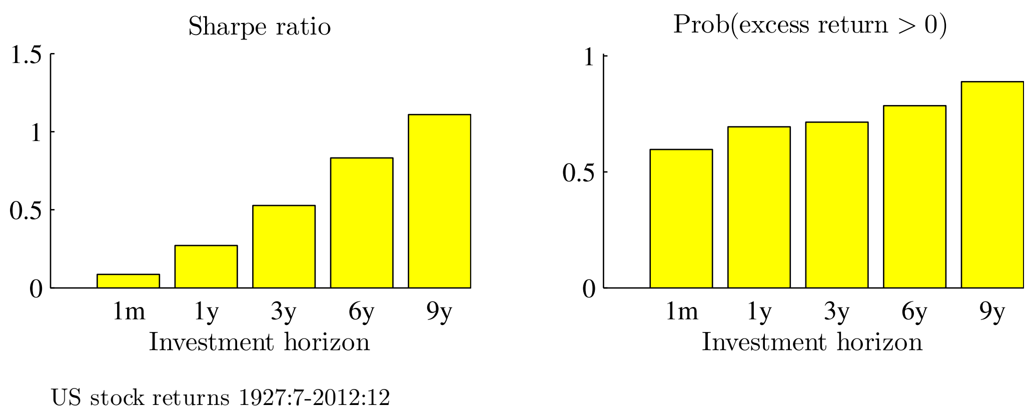


Figure 9.1: SR and probability of excess return>0

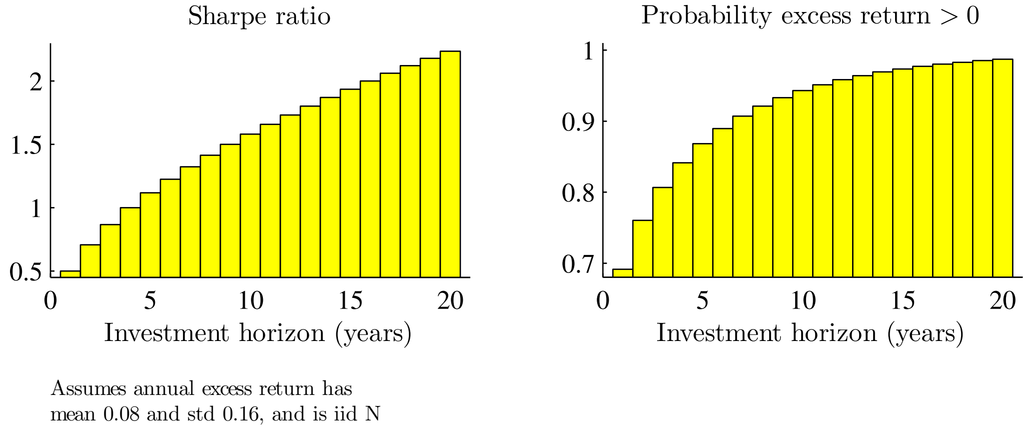


Figure 9.2: SR and probability of excess return>0, iid returns

the square root of horizon. However, this does *not* mean that risky assets are better for long horizons, at least not if we believe in mean variance preferences and unpredictable returns. Something else than iid data is needed for that.

Let Z_q be the net return on a q-period investment. If returns are iid, the Sharpe ratio of Z_q is approximately

$$SR(Z_q) \approx \sqrt{q}\frac{\mathrm{E}\, R^e}{\mathrm{Std}(R)}, \tag{9.1}$$

where $\mathrm{E}\, R^e$ is the mean one-period excess return and $\mathrm{Std}(R)$ is the standard deviation of the one-period return. (Time subscripts are suppressed to keep the notation simple.) This Sharpe ratio is clearly increasing with the horizon, q.

Proof. (of (9.1)) The q-period net return is

$$\begin{aligned} Z_q &= (R_1 + 1)(R_2 + 1)\dots(R_q + 1) - 1 \\ &\approx R_1 + R_2 + \dots + R_q. \end{aligned}$$

If returns are iid, then the mean and variance of the q-period return are approximately

$$\begin{aligned} \mathrm{E}\, Z_q &\approx q\, \mathrm{E}\, R, \\ \mathrm{Var}(Z_q) &\approx q\, \mathrm{Var}(R). \end{aligned}$$

■

Example 9.1 *(The quality of the approximation of the q-period return) If $R_1 = 0.9$ and $R_2 = -0.9$, then the two-period net return is*

$$Z_2 = (1 + 0.9)(1 - 0.9) - 1 = -0.81$$

With the approximation we instead have

$$Z_2 \approx R_1 + R_2 = 0.$$

The difference in net returns is dramatic. If the two net returns instead are $R_1 = 0.09$ and $R_2 = -0.09$, then

$$Z_2 = (1 + 0.09)(1 - 0.09) - 1 = -0.01$$

and the approximation is still zero: the difference is much smaller.

Example 9.2 *(The danger of arithmetic mean return). Consider two portfolios with the following returns*

	Portfolio A	*Portfolio B*
Year 1	5%	20%
Year 2	−5%	−35%
Year 3	5%	25%

Just adding these returns give 5% and 10% respectively, but the total returns over the three periods are actually 4.7% and -2.5% respectively.

9.1.2 Probability of Outperforming a Riskfree Asset

Since the Sharpe ratio is increasing with the investment horizon, the probability of beating a riskfree asset is (typically) also increasing. To simplify, assume that the returns are normally distributed. Then, we have

$$\Pr\left(Z_q^e > 0\right) = \Phi\left[SR(Z_q)\right], \tag{9.2}$$

where Z_q^e is the excess return on a q-period investment and $\Phi()$ is the cumulative distribution function of a standard normal variable, $N(0, 1)$. The argument of an increasing probability of a positive excess return is therefore the same argument as the increasing Sharpe ratio. See Figure 9.2 for an illustration.

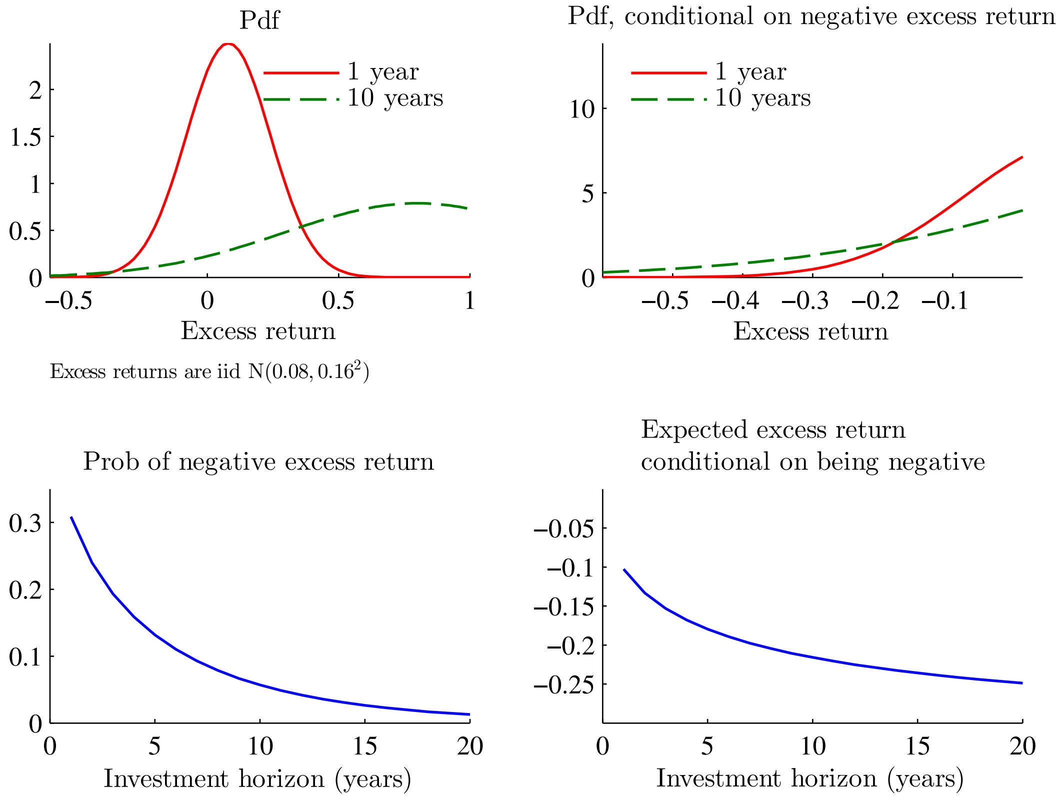


Figure 9.3: Time diversification, normally distributed returns

Proof. (of (9.2)) By standard manipulations we have

$$
\begin{aligned}
\Pr\left(Z_q^e > 0\right) &= 1 - \Pr\left(Z_q^e \leq 0\right) \\
&= 1 - \Pr\left(\frac{Z_q^e - \mathrm{E}\, Z_q^e}{\operatorname{Std}(Z_q^e)} \leq \frac{-\mathrm{E}\, Z_q^e}{\operatorname{Std}(Z_q^e)}\right) \\
&= 1 - \Phi\left(\frac{-\mathrm{E}\, Z_q^e}{\operatorname{Std}(Z_q^e)}\right) \\
&= \Phi\left(\frac{\mathrm{E}\, Z_q^e}{\operatorname{Std}(Z_q^e)}\right),
\end{aligned}
$$

where the last line follows from $\Phi(x)+\Phi(-x) = 1$ since the standard normal distribution is symmetric around zero. ■

9.1.3 MV Portfolio Choice

Although the increasing Sharpe ratios mean that the probability of beating a riskfree asset is increasing with the investment horizon, that does not mean that the risky asset is safer for a long-run investor. The reason is, of course, that we also have to take into account the size of the loss—in case the portfolio underperforms. With a longer horizon (and therefore higher dispersion), really bad outcomes are more likely—so the expected loss (conditional of having one) is increasing with the investment horizon. See Figure 9.3 for an illustration.

Remark 9.3 *(Expected excess return conditional on a negative one*) If $x \sim N(\mu, \sigma^2)$, then $\mathrm{E}(x|x \leq b) = \mu - \sigma\phi(b_0)/\Phi(b_0)$ where $b_0 = (b - \mu)/\sigma$ and where $\phi()$ and $\Phi()$ are the pdf and cdf of a $N(0,1)$ variable respectively. To apply this, use $b = 0$ so $b_0 = -\mu/\sigma$. This gives $\mathrm{E}(x|x \leq 0) = \mu - \sigma\phi(-\mu/\sigma)/\Phi(-\mu/\sigma)$. Here this gives*

$$\mathrm{E}(Z_q^e|Z_q^e \leq 0) = \mathrm{E}\, Z_q^e - \mathrm{Std}(Z_q^e)\frac{\phi[-SR(Z_q)]}{\Phi[-SR(Z_q)]},$$

which for iid returns equals

$$\mathrm{E}(Z_q^e|Z_q^e \leq 0) = q\,\mathrm{E}\, Z_1^e - \sqrt{q}\,\mathrm{Std}(Z_1^e)\frac{\phi[-\sqrt{q}SR(Z_1)]}{\Phi[-\sqrt{q}SR(Z_1)]}.$$

For most reasonable values (for equity markets), this is decreasing in q. (Actually, numerical calculations suggests that it is always decreasing in q, but I have no formal proof (yet)).

To say more about how the investment horizon affects the portfolio weights, we need to be more precise about the preferences. As a benchmark, consider a mean-variance investor who will choose a portfolio for q periods. With one risky asset (the tangency portfolio) and a riskfree asset, the optimization problem is

$$\max_v v\,\mathrm{E}\, Z_q^e + qR_f - \frac{k}{2}v^2\,\mathrm{Var}(Z_q), \tag{9.3}$$

where R_f is the per-period riskfree rate. With iid returns, both the mean and the variance scale linearly with the investment horizon, so we can equally well write the optimization problem as

$$\max_v vq\,\mathrm{E}\, R^e + qR_f - \frac{k}{2}v^2 q\,\mathrm{Var}(R), \text{ if iid returns.} \tag{9.4}$$

Clearly, scaling this objective function by $1/q$ will not change anything: the horizon is irrelevant.

To be more precise, the solution of (9.3) is

$$v = \frac{1}{k} \frac{\mathrm{E}\, Z_q^e}{\mathrm{Var}(Z_q)}. \tag{9.5}$$

If returns are iid, we get the following portfolio weights for investment horizons of one and two periods

$$v(1) = \frac{1}{k} \frac{\mathrm{E}\, R^e}{\mathrm{Var}(R)}, \tag{9.6}$$

$$v(2) = \frac{1}{k} \frac{2\, \mathrm{E}\, R^e}{2\, \mathrm{Var}(R)}, \tag{9.7}$$

which are the same. With MV behaviour, *non-iid returns are required to generate a horizon effect on the portfolio choice*. The key point is that the portfolio weight is not determined by the Sharpe ratio, but the Sharpe ratio divided by the standard deviation. Or to put it another way, comparing Sharpe ratios across investment horizons is not very informative.

Proof. (of (9.5)) The first order condition of (9.3) is

$$0 = \mathrm{E}\, Z_q^e - kv\, \mathrm{Var}(Z_q) \text{ or}$$
$$v = \frac{1}{k} \frac{\mathrm{E}\, Z_q^e}{\mathrm{Var}(Z_q)}.$$

■

Example 9.4 *(US long-run stock market) For the period 1947–2001, the US stock market had an average excess return of 8% (per year) and a standard deviation of 16%. From (9.5), the weight on the risky asset is then* $v = (0.08/0.16^2)/k = 3.125/k$.

With autocorrelated returns two things change: returns are predictable so the expected return is time-varying, and the variance of the two-period return includes a covariance term. The portfolio weights (chosen in period 0) are then

$$v(1) = \frac{1}{k}\frac{\mathrm{E}_0\, R^e_1}{\mathrm{Var}_0(R_1)}, \tag{9.8}$$

$$v(2) = \frac{1}{k}\frac{\mathrm{E}_0(R^e_1 + R^e_2)}{\mathrm{Var}_0(R_1) + \mathrm{Var}_0(R_2) + 2\,\mathrm{Cov}_0(R_1, R_2)}, \tag{9.9}$$

where all moments carry a time subscript to indicate that they are conditional moments. A key aspect of these formulas is that mean reversion in prices makes the covariance (of returns) negative. This will tend to make the weight for the two-period horizon larger. The intuition is simple: with mean reversion in prices, long-run investments are less risky than short-run investments since extreme movements will be partially "averaged out" over time. Empirically, there is some evidence of mean-reversion on the business cycle frequencies (a couple of years). The effect is not strong, however, so mean reversion is probably a poor argument for horizon effects.

Example 9.5 *(AR(1) process for returns) Suppose the excess returns follow an AR(1) process*

$$R^e_{t+1} = \mu(1-\rho) + \rho R^e_t + \varepsilon_{t+1} \text{ with } \sigma^2 = \mathrm{Var}(\varepsilon_{t+1}).$$

The conditional moments are then

$$\begin{aligned}
\mathrm{E}_0\, R^e_1 &= \mu(1-\rho) + \rho R^e_0,\\
\mathrm{E}_0\, R^e_2 &= \mu(1-\rho^2) + \rho^2 R^e_0,\\
\mathrm{Var}_0(R_1) &= \sigma^2\\
\mathrm{Var}_0(R_2) &= (1+\rho^2)\sigma^2\\
\mathrm{Cov}_0(R_1, R_2) &= \rho\sigma^2.
\end{aligned}$$

If the initial return is at the mean, $R^e_0 = \mu$, then the forecasted return is μ across all horizons, which gives the portfolio weights

$$\begin{aligned}
v(1) &= \frac{1}{k}\frac{\mu}{\sigma^2},\\
v(2) &= \frac{1}{k}\frac{\mu}{\sigma^2}\frac{2}{(2+\rho^2+2\rho)}.
\end{aligned}$$

With $\rho = (-0.5, 0, 0.5)$ *the last term is around* $(1.6, 1, 0.6)$. *With* $\rho = (-0.1, 0, 0.1)$*, the last term is around* $(1.1, 1, 0.9)$.

9.2 Time Diversification and the Growth-Optimal Portfolio: Lognormal Returns

This section revisits the issue of time diversification—this time in a setting where log portfolio returns are normally distributed. This allows us to get more precise results, since we can avoid approximating the cumulative returns.

9.2.1 Time Diversification with Lognormal Returns

The gross return on a q-period investment can be written

$$1 + Z_q = (1 + R_1)(1 + R_2)...(1 + R_q), \tag{9.10}$$

where R_t is the net portfolio return in period t. Taking logs (and using lower case letters to denote them), we have the log q-period return

$$z_q = r_1 + r_2 + \ldots + r_q, \tag{9.11}$$

where $z_q = \ln(1 + Z_q)$ and $r_t = \ln(1 + R_t)$.

Remark 9.6 *($\ln(1 + x) \approx x$...) If x is small, $\ln(1 + x) \approx x$, so assuming that x is normally distributed is fairly similar to assuming that $\ln(1 + x)$ is normally distributed.*

Remark 9.7 *(Lognormal distribution) If $x \sim N(\mu, \sigma^2)$ and $y = \exp(x)$, then the probability density function of y is*

$$\text{pdf}(y) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{\ln y - \mu}{\sigma}\right)^2\right], \; y > 0.$$

The rth moment of y is $\mathrm{E}\, y^r = \exp(r\mu + r^2\sigma^2/2)$.

To simplify the analysis, assume that the log returns of portfolio y, r_{yt}, are iid $N(\mu_y, \sigma_y^2)$. (This is a convenient assumption since it carries over to multi-period returns.) The "Sharpe ratio" of the log q-period return, z_{qy}, is

$$SR(z_{qy}) = \sqrt{q}\frac{\mu_y - r_f}{\sigma_y}, \tag{9.12}$$

where r_f is the continuously compounded interest rate.

If log returns are normally distributed, the probability of the q-period return of portfolio y (denoted Z_{qy}) being higher than the q-return of portfolio x (Z_{qx}) is

$$\Pr\left(Z_{qy} > Z_{qx}\right) = \Phi\left(\sqrt{q}\frac{\mu_y - \mu_x}{\sigma\left(r_y - r_x\right)}\right), \tag{9.13}$$

where Φ is the cumulative distribution function of a standard normal variable, $N(0, 1)$, μ_y the expected log return on portfolio y, and $\sigma\left(r_y - r_x\right)$ is the standard deviation of the difference in log returns. (The portfolios are constant over time, since the returns are iid.) In particular, if the x portfolio is a riskfree asset with log return r_f, then the probability is

$$\Pr\left(Z^e_{qy} > 0\right) = \Phi\left[SR(z_{qy})\right], \tag{9.14}$$

which is a function of the Sharpe ratio for the log returns. This probability is clearly increasing with the investment horizon, q. On the other hand, with a longer horizon (and therefore higher dispersion), really bad outcomes more likely.

See Figure 9.4 for an illustration.

Proof. (of (9.12)) Consider (9.11). If log returns are iid with mean μ and variance σ^2, then the mean and variance of the q-period return are

$$\begin{aligned}
\mathrm{E}\, z_q &= q\mu, \\
\mathrm{Var}(z_q) &= q\sigma^2.
\end{aligned}$$

■

Proof. (*of (9.13)) By standard manipulations we have

$$\begin{aligned}
\Pr\left[\exp\left(\textstyle\sum_{t=1}^q r_{ty}\right) > \exp\left(\textstyle\sum_{t=1}^q r_{tx}\right)\right] &= 1 - \Pr\left[\exp\left(\textstyle\sum_{t=1}^q r_{ty}\right) \leq \exp\left(\textstyle\sum_{t=1}^q r_{tx}\right)\right] \\
&= 1 - \Pr\left[\textstyle\sum_{t=1}^q r_{ty} \leq \sum_{t=1}^q r_{tx}\right] \\
&= 1 - \Pr\left[\frac{\sum_{t=1}^q \left(r_{ty} - r_{tx}\right) - q\left(\mu_y - \mu_x\right)}{\sqrt{q}\sigma\left(r_{yt} - r_{xt}\right)} \leq -\frac{q\left(\mu_y - \mu_x\right)}{\sqrt{q}\sigma\left(r_{yt} - r_{xt}\right)}\right] \\
&= 1 - \Phi\left[-\sqrt{q}\frac{\mu_y - \mu_x}{\sigma\left(r_{yt} - r_{xt}\right)}\right] \\
&= \Phi\left[\sqrt{q}\frac{\mu_y - \mu_x}{\sigma\left(r_{yt} - r_{xt}\right)}\right],
\end{aligned}$$

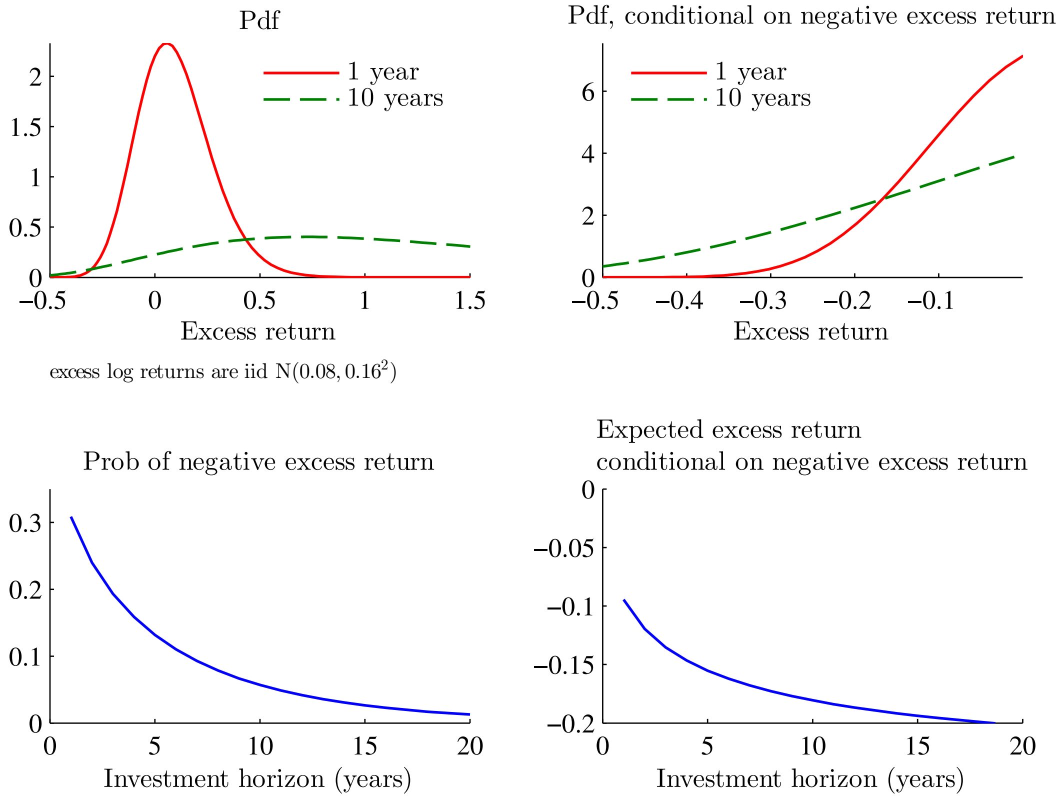


Figure 9.4: Time diversification, lognormally distributed returns

where the last line follows from $\Phi(z)+\Phi(-z) = 1$ since the standard normal distribution is symmetric around zero. ■

Remark 9.8 *(Expected excess return conditional on a negative one*) If* $\ln y \sim N(\mu, \sigma^2)$, *then* $\mathrm{E}(y|y \leq b) = \exp\left(\mu + \sigma^2/2\right) \Phi(-\sigma + b_0)/\Phi(b_0)$ *with* $b_0 = (\ln b - \mu)/\sigma$, *where* $\Phi()$ *is the cdf of a* $N(0,1)$ *variable. To apply this, use* $\ln b = 0$ *so* $b_0 = -\mu/\sigma$. *We then have* $\mathrm{E}(y|y \leq 1) = \exp\left(\mu + \sigma^2/2\right) \Phi(-\sigma - \mu/\sigma)/\Phi(-\mu/\sigma)$. *Here this gives that the expected gross return of the riskfree asset, divided by the gross return of the riskfree asset is*

$$
\begin{aligned}
&= \exp\left(\mu + \sigma^2/2\right) \Phi(-\sigma - SR_q)/\Phi(-SR_q) \\
&= \exp[q(\mu + \sigma^2/2)]\Phi[(-\sqrt{q}(\sigma + SR_q)]/\Phi(-\sqrt{q}SR_q),
\end{aligned}
$$

where the second line is for iid returns.

9.2.2 Portfolio Choice with a Logarithmic Utility Function

To demonstrate that, with iid log returns, optimal portfolio weights are indeed unaffected by the investment horizon, consider the simple case of a logarithmic utility function, where we find a portfolio that solves

$$\max_v \mathrm{E} \ln(1 + R_q) = \max_v \mathrm{E}(r_1 + r_2 + \ldots + r_q), \tag{9.15}$$

where r_t is the log portfolio return in period t (which clearly depends on the chosen portfolio weights v). We here assume that the portfolio weights are chosen at the beginning (time $t = 0$) of the investment period and then kept unchanged. With iid log returns, we can clearly write (9.15) as

$$\max_v q \, \mathrm{E} \, r_1, \tag{9.16}$$

which demonstrates that the investment horizon does not matter for the optimal portfolio choice. It doesn't matter that the Sharpe ratio is increasing.

Example 9.9 *(Portfolio choice with logarithmic utility function) It is typically hard to find explicit expressions for what the portfolio weights should be with log utility, so one typically has to resort to numerical methods. This example shows a case where we can find an explicit solution—because of a very simple setting. Suppose there are two states (1 and 2) and that asset A has the gross return $R_A(1)$ in state 1 and $R_A(2)$ in state 2—and similarly for asset B. The portfolio return is $R_p = vR^e + R_B$, where $R^e = R_A - R_B$. If π is the probability of state 1, then the expected log portfolio return is*

$$\mathrm{E} \ln(R_p) = \pi \ln[vR^e(1) + R_B(1)] + (1 - \pi) \ln[vR^e(2) + R_B(2)].$$

The first order condition for v is

$$0 = \frac{\pi}{vR^e(1) + R_B(1)} R^e(1) + \frac{(1 - \pi)}{vR^e(2) + R_B(2)} R^e(2)$$

and the solution is

$$v = -\frac{\pi R^e(1) R_B(2) + (1 - \pi) R^e(2) R_B(1)}{R^e(1) R^e(2)}.$$

See Figure 9.5 for an illustration.

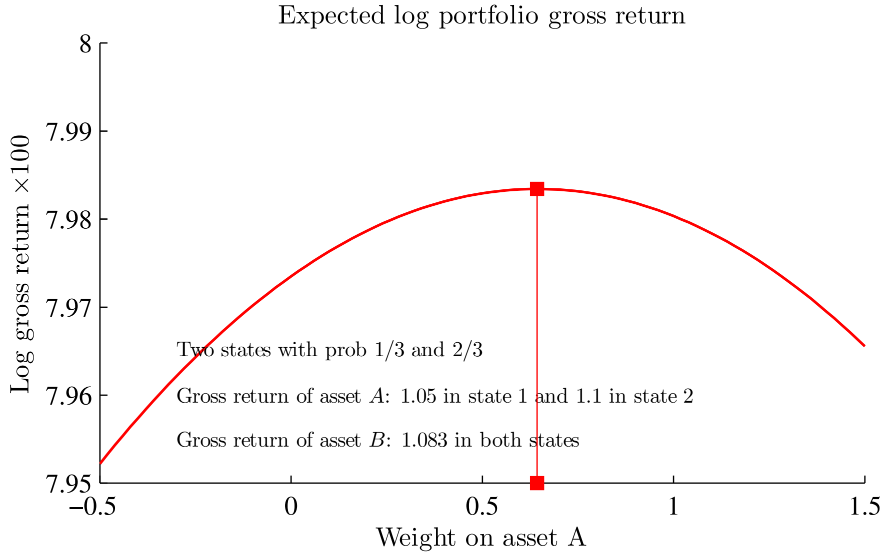


Figure 9.5: Example of portfolio choice with log utility

Remark 9.10 *(Comparison of geometric and arithmetic mean returns*) Let S_t be the asset price in period t. The geometric mean return g satisfies*

$$\left(S_q/S_0\right)^{1/q} = 1 + g$$

so the log can be written

$$\ln(1+g) = \frac{1}{q}\ln(S_q/S_0) = \frac{1}{q}\left(r_1 + r_2 + \ldots + r_q\right),$$

where $r_t = \ln(1+R_t)$ is the log return. If μ is the average log return, then the expected value is

$$\mathrm{E}\ln(1+g) = \frac{1}{q}\textstyle\sum_{t=1}^{q}\mathrm{E}\,r_t = \mu.$$

The arithmetic mean return is defined as

$$R_{arithmetic} = \frac{1}{q}\left(R_1 + R_2 + \ldots + R_q\right).$$

If r_t is iid $N(\mu,\sigma^2)$, then we get

$$\mathrm{E}\,R_{arithmetic} = \frac{1}{q}\textstyle\sum_{t=1}^{q}\exp\left(\mu+\sigma^2/2\right) = \exp\left(\mu+\sigma^2/2\right).$$

To make it comparable with the geometric mean return, take logs to get

$$\ln \mathrm{E}\, R_{arithmetic} = \mu + \sigma^2/2.$$

Hence, we have that (for log returns)

$$\textit{arithmetic mean return} = \textit{geometric mean return} + \sigma^2/2.$$

Clearly, they coincide when the returns are constant over time.

Example 9.11 *(Arithmetic and geometric mean returns) Consider the following table*

	Portfolio A	*Portfolio B*
Year 1	5%	20%
Year 2	–5%	–35%
Year 3	5%	25%
Arithmetic mean	1.67%	3.33%
Geometric mean	1.55%	–0.84%

In this case, the rankings (of the portfolios) based the arithmetic and geometric means are different.

9.2.3 The Growth-Optimal Portfolio and Log Utility

The portfolio that comes out from maximizing the log return has some interesting properties. If portfolio y has the highest expected log return, then (9.13) shows that the probability that it beats any other portfolio is increasing with the investment horizon—and goes to unity as the horizon goes to infinity. This portfolio is called the *growth-optimal portfolio*.

See Figure 9.6 for an illustration.

This portfolio is commonly advocated to be the best for any long-run investor. That argument is clearly flawed. In particular, for an investor with a relative risk aversion different from one, the growth-optimal portfolio is *not* optimal: a higher risk aversion would give a more conservative portfolio. (It can be shown that the logarithmic utility function is a CRRA utility function with a relative risk aversion of one.) The intuition is

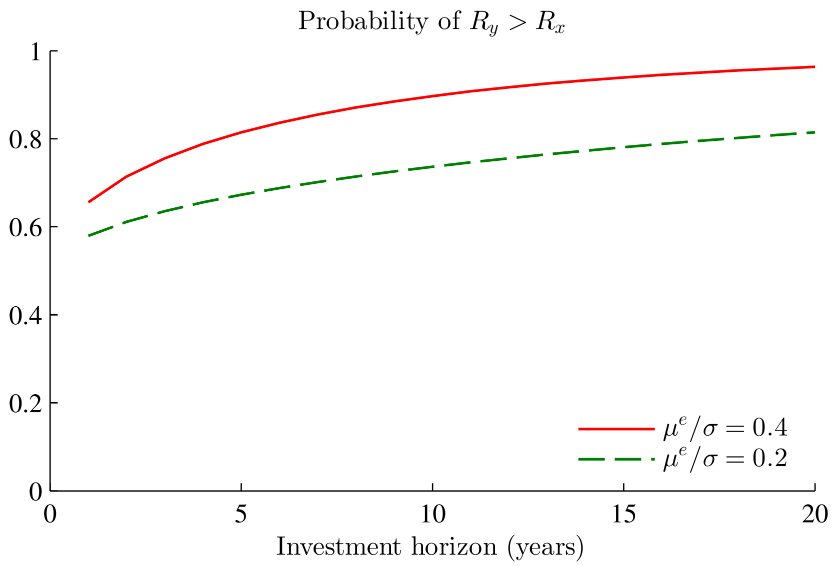


Figure 9.6: The probability of outperforming another portfolio

that the occasional lower return of the growth-optimal portfolio is considered very risky, so the investor prefers a less volatile portfolio.

Notice that, for a given $q < \infty$, the growth-optimal portfolio does not necessarily maximize the probability of beating other portfolios. While the growth-optimal portfolio has the highest expected log return so it maximizes the numerator in (9.13), it may well have a very high volatility. It is only in the limit that the growth-optimal portfolio is a sure winner.

9.2.4 Maximizing the Geometric Mean Return

The growth-optimal portfolio is often said to maximize the geometric mean return. That is true, but may need a clarification.

Remark 9.12 *(Geometric mean) Suppose the random variable x can take the values $x(1), x(2), \ldots, x(S)$ with probabilities $\pi(1), \pi(2), \ldots, \pi(S)$, where $\sum_{j=1}^{S} \pi(j) = 1$. The arithmetic mean (expected value) is $\sum_{j=1}^{S} \pi(j) x(j)$ and the geometric mean is $\prod_{j=1}^{S} x(j)^{\pi(j)}$. Taking the log of the definition of a geometric mean gives*

$$\sum_{j=1}^{S} \pi(j) \ln x(j) = \mathrm{E} \ln x,$$

which is the expected value of the log of x.

Remark 9.13 *(Sample geometric mean) With the sample $z_1, z_2, \ldots, z_T$, the sample arithmetic mean is $\sum_{t=1}^{T} z_t / T$ and the sample geometric mean is $\prod_{t=1}^{T} z_t^{1/T}$.*

It follows directly from these remarks that a portfolio that maximizes the geometric mean of the portfolio gross return $1 + R_p$ also maximizes the expected log return of it, $\mathrm{E} \ln \left(1 + R_p\right)$.

An intuitive way of motivating this portfolio is as follows. The gross return on the q-period investment in (9.10) is, of course, random, but in a very large sample (long investment horizon), the histogram of the returns should start to converge to the true distribution. With iid returns, this is the same distribution that defined the geometric mean (which we have maximized). Hence, with a very long investment period, the portfolio (that maximizes the geometric mean) should give the highest return over the investment period. Of course, this is virtually the same argument as in (9.13), which showed that the growth-optimal portfolio will outperform all other portfolios with probability one as the investment horizon goes to infinity. (The only difference is that the current argument does not rely on the normal distribution of the log returns.)

9.3 More General Utility Functions and Rebalancing

We will now take a look at more general optimization problems. Assume that the objective is to maximize

$$\mathrm{E}_0 \, u(W_q), \tag{9.17}$$

where W_q is the wealth (in real terms) at time q (the investment horizon) and E_0 denotes the expectations formed in period 0 (the initial period). What can be said about how the investment horizon affects the portfolio weights?

If the investor is not allowed (or it is too costly) to rebalance the portfolio—and the utility function/distribution of returns are such that the investor picks a mean-variance portfolio (quadratic utility function or normally distributed returns), then the results in Section 9.1.1 go through: non-iid returns are required to generate a horizon effect on the portfolio choice.

If, more realistically, the investor is allowed to rebalance the portfolio, then the analysis is more difficult. We summarize some known results below.

9.3.1 CRRA Utility Function and iid Returns

Suppose the utility function has constant relative risk aversion, so the objective in period 0 is

$$\max \mathrm{E}_0 \, W_q^{1-\gamma}/(1-\gamma). \tag{9.18}$$

In period one, the objective is $\max \mathrm{E}_1 \, W_q^{1-\gamma}/(1-\gamma)$, which may differ in terms of what we know about the distribution of future returns (incorporated into the expectations operator) and also in terms of the current wealth level (due to the return in period 1).

With CRRA utility, relative portfolio weights are independent of the wealth of the investor (fairly straightforward to show). If we combine this with iid returns—then the only difference between an investor in t and the same investor in $t+1$ is that he may be poorer or wealthier. This investor will therefore choose the same portfolio weights in every period. Analogously, a short run investor and a long run investor choose the same portfolio weights (you can think of the investor in $t+1$ as a short run investor). Therefore, with a CRRA utility function and iid returns there are no horizon effects on the portfolio choice. In addition, the portfolio weights will stay constant over time. The intuition is that all periods look the same.

However, with non-iid returns (predictability or variations in volatility) there will be horizon effects (and changes in weights over time). This would give rise to *intertemporal hedging*, where the choice of today's portfolio is affected by the likely changes of the investment opportunities tomorrow.

The same result holds if the objective function instead is to maximize the utility from stream of consumption, provided the utility function is CRRA and time separable. In this case, the objective is

$$\max C_0^{1-\gamma}/(1-\gamma) + \delta \, \mathrm{E}_0 \, C_1^{1-\gamma}/(1-\gamma) + \ldots + \delta^q \, \mathrm{E}_0 \, C_q^{1-\gamma}/(1-\gamma). \tag{9.19}$$

The basic mechanism is that the optimal consumption/wealth ratio turns out to be constant.

9.3.2 Logarithmic Utility Function and non-iid Returns

In the special case where the relative risk aversion (in a CRRA utility function) is one, then the utility function becomes logarithmic.

The objective in period 0 is then

$$\max \mathrm{E}_0 \ln W_q = \max(\ln W_0 + \mathrm{E}_0 r_1 + \mathrm{E}_0 r_2 + \ldots + \mathrm{E}_0 r_q), \tag{9.20}$$

where r_t is the log return, $r_t = \ln(1 + R_t)$ where R_t is a net return.

Since the returns in the different periods enter separably, the best an investor can do in period 0 is to choose a portfolio that maximizes $\mathrm{E}_0 r_1$—that is, to choose the one-period growth-optimal portfolio. But, a short run investor who maximizes $\mathrm{E}_0 \ln[W_0(1 + R_1)]) = \max(\ln W_0 + \mathrm{E}_0 r_1)$ will choose the same portfolio. There is then no horizon effect. However, the portfolio choice may change over time, if the distribution of the returns do.

The same result holds if the objective function instead is to maximize the utility from stream of consumption as in (9.19), but with a logarithmic utility function.

Bibliography

Campbell, J. Y., and L. M. Viceira, 2002, *Strategic asset allocation: portfolio choice of long-term investors*, Oxford University Press.

Elton, E. J., M. J. Gruber, S. J. Brown, and W. N. Goetzmann, 2010, *Modern portfolio theory and investment analysis*, John Wiley and Sons, 8th edn.

10 Efficient Markets

Reference (medium): Elton, Gruber, Brown, and Goetzmann (2010) 17 (efficient markets) and 26 (earnings estimation)

Additional references: Campbell, Lo, and MacKinlay (1997) 2 and 7; Cochrane (2001) 20.1

More advanced material is denoted by a star (*). It is not required reading.

10.1 Asset Prices, Random Walks, and the Efficient Market Hypothesis

Let P_t be the price of an asset at the end of period t, after any dividend in t has been paid (an ex-dividend price). The gross return ($1 + R_{t+1}$, like 1.05) of holding an asset with dividends (per current share), D_{t+1}, between t and $t + 1$ is then defined as

$$1 + R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t}. \tag{10.1}$$

The dividend can, of course, be zero in a particular period, so this formulation encompasses the case of daily stock prices with annual dividend payment.

Remark 10.1 *(Conditional expectations) The expected value of the random variable y_{t+1} conditional on the information set in t, $\mathrm{E}_t\, y_{t+1}$ is the best guess of y_{t+1} using the information in t. Example: suppose y_{t+1} equals $x_t + \varepsilon_{t+1}$, where x_t is known in t, but all we know about ε_{t+1} in t is that it is a random variable with a zero mean and some (finite) variance. In this case, the best guess of y_{t+1} based on what we know in t is equal to x_t.*

Take expectations of (10.1) based on the information set in t

$$1 + \mathrm{E}_t\, R_{t+1} = \frac{\mathrm{E}_t\, P_{t+1} + \mathrm{E}_t\, D_{t+1}}{P_t} \text{ or} \tag{10.2}$$

$$P_t = \frac{\mathrm{E}_t P_{t+1} + \mathrm{E}_t D_{t+1}}{1 + \mathrm{E}_t R_{t+1}}. \tag{10.3}$$

This formulation is only a definition, but it will help us organize the discussion of how asset prices are determined.

This expected return, $\mathrm{E}_t\, R_{t+1}$, is likely to be greater than a riskfree interest rate if the asset has positive systematic (non-diversifiable) risk. For instance, in a CAPM model this would manifest itself in a positive "beta." In an equilibrium setting, we can think of this as a "required return" needed for investors to hold this asset.

10.1.1 Different Versions of the Efficient Market Hypothesis

The efficient market hypothesis casts a long shadow on every attempt to forecast asset prices. In its simplest form it says that it is not possible to forecast asset prices, but there are several other forms with different implications. Before attempting to forecast financial markets, it is useful to take a look at the logic of the efficient market hypothesis. This will help us to organize the effort and to interpret the results.

A *modern interpretation of the efficient market hypothesis* (EMH) is that the information set used in forming the market expectations in (10.2) includes all public information. (This is the semi-strong form of the EMH since it says all public information; the strong form says all public and private information; and the weak form says all information in price and trading volume data.) The implication is that simple stock picking techniques are not likely to improve the portfolio performance, that is, abnormal returns. Instead, advanced (costly?) techniques are called for in order to gather more detailed information than that used in market's assessment of the asset. Clearly, with a better forecast of the future return than that of the market there is plenty of scope for dynamic trading strategies. Note that this modern interpretation of the efficient market hypothesis does not rule out the possibility of forecastable prices or returns. It does rule out that abnormal returns can be achieved by stock picking techniques which rely on public information.

There are several different *traditional interpretations of the EMH*. Like the modern interpretation, they do not rule out the possibility of achieving abnormal returns by using better information than the rest of the market. However, they make stronger assumptions about whether prices or returns are forecastable. Typically one of the following is assumed to be unforecastable: price changes, returns, or returns in excess of a riskfree rate (interest rate). By unforecastable, it is meant that the best forecast (expected value conditional on available information) is a constant. Conversely, if it is found that there is some information in t that can predict returns R_{t+1}, then the market cannot price the asset as if $\mathrm{E}_t\, R_{t+1}$ is a constant—at least not if the market forms expectations rationally. We will now analyze the logic of each of the traditional interpretations.

If price changes are unforecastable, then $\mathrm{E}_t\, P_{t+1} - P_t$ equals a constant. Typically, this constant is taken to be zero so P_t is a martingale. Use $\mathrm{E}_t\, P_{t+1} = P_t$ in (10.2)

$$\mathrm{E}_t\, R_{t+1} = \frac{\mathrm{E}_t\, D_{t+1}}{P_t}. \tag{10.4}$$

This says that the expected net return on the asset is the expected dividend divided by the current price. This is clearly implausible for daily data since it means that the expected return is zero for all days except those days when the asset pays a dividend (or rather, the day the asset goes ex dividend)—and then there is an enormous expected return for the one day when the dividend is paid. As a first step, we should probably refine the interpretation of the efficient market hypothesis to include the dividend so that $\mathrm{E}_t(P_{t+1} + D_{t+1}) = P_t$. Using that in (10.2) gives $1 + \mathrm{E}_t\, R_{t+1} = 1$, which can only be satisfied if $\mathrm{E}_t\, R_{t+1} = 0$, which seems very implausible for long investment horizons—although it is probably a reasonable approximation for short horizons (a week or less).

If returns are unforecastable, so $\mathrm{E}_t\, R_{t+1} = R$ (a constant), then (10.3) gives

$$P_t = \frac{\mathrm{E}_t P_{t+1} + \mathrm{E}_t D_{t+1}}{1 + R}. \tag{10.5}$$

The main problem with this interpretation is that it looks at every asset separately and that outside options are not taken into account. For instance, if the nominal interest rate changes from 5% to 10%, why should the expected (required) return on a stock be unchanged? In fact, most asset pricing models suggest that the expected return $\mathrm{E}_t\, R_{t+1}$ equals the riskfree rate plus compensation for risk.

If excess returns are unforecastable, then the compensation (over the riskfree rate) for risk is constant. The risk compensation is, of course, already reflected in the current price P_t, so the issue is then if there is some information in t which is correlated with the risk compensation in P_{t+1}. Note that such predictability does not necessarily imply an inefficient market or presence of uninformed traders—it could equally well be due to movements in risk compensation driven by movements in uncertainty (option prices suggest that there are plenty of movements in uncertainty). If so, the predictability cannot be used to generate abnormal returns (over riskfree rate plus risk compensation). However, it could also be due to exploitable market inefficiencies. Alternatively, you may argue that the market compensates for risk which you happen to be immune to—so you are interested in the return rather than the risk adjusted return.

This discussion of the traditional efficient market hypothesis suggests that the most interesting hypotheses to test are if returns or excess returns are forecastable. In practice, the results for them are fairly similar since the movements in most asset returns are much greater than the movements in interest rates.

10.1.2 Martingales and Random Walks*

Further reading: Cuthbertson (1996) 5.3

The accumulated wealth in a sequence of fair bets is expected to be unchanged. It is then said to be a martingale.

The time series x is a martingale with respect to an information set Ω_t if the expected value of x_{t+s} ($s \geq 1$) conditional on the information set Ω_t equals x_t. (The information set Ω_t is often taken to be just the history of x: $x_t, x_{t-1}, ...$)

The time series x is a random walk if $x_{t+1} = x_t + \varepsilon_{t+1}$, where ε_t and ε_{t+s} are uncorrelated for all $s \neq 0$, and $\mathrm{E}\,\varepsilon_t = 0$. (There are other definitions which require that ε_t and ε_{t+s} have the same distribution.) A random walk is a martingale; the converse is not necessarily true.

Remark 10.2 *(A martingale, but not a random walk). Suppose $y_{t+1} = y_t u_{t+1}$, where u_t and u_{t+s} are uncorrelated for all $s \neq 0$, and $\mathrm{E}_t\, u_{t+1} = 1$. This is a martingale, but not a random walk.*

In any case, the martingale property implies that $x_{t+s} = x_t + \varepsilon_{t+s}$, where the expected value of ε_{t+s} based on Ω_t is zero. This is close enough to the random walk to motivate the random walk idea in most cases.

10.2 Autocorrelations

10.2.1 Autocorrelation Coefficients

The autocovariances of the y_t process can be estimated as

$$\hat{\gamma}_s = \frac{1}{T} \sum_{t=1+s}^{T} (y_t - \bar{y})(y_{t-s} - \bar{y}), \text{ with} \tag{10.6}$$

$$\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t. \tag{10.7}$$

(We typically divide by T in (10.6) even if we have only $T - s$ full observations to estimate γ_s from.) Autocorrelations are then estimated as

$$\hat{\rho}_s = \hat{\gamma}_s / \hat{\gamma}_0. \tag{10.8}$$

The sampling properties of $\hat{\rho}_s$ are complicated, but there are several useful large sample results for Gaussian processes (these results typically carry over to processes which are similar to the Gaussian—a homoskedastic process with finite 6th moment is typically enough, see Priestley (1981) 5.3 or Brockwell and Davis (1991) 7.2-7.3). When the true autocorrelations are all zero (not ρ_0, of course), then for any i and j different from zero

$$\sqrt{T} \begin{bmatrix} \hat{\rho}_i \\ \hat{\rho}_j \end{bmatrix} \to^d N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right). \tag{10.9}$$

This result can be used to construct tests for both single autocorrelations (t-test or χ^2 test) and several autocorrelations at once (χ^2 test).

Example 10.3 *(t-test) We want to test the hypothesis that* $\rho_1 = 0$. *Since the* $N(0,1)$ *distribution has 5% of the probability mass below -1.65 and another 5% above 1.65, we can reject the null hypothesis at the 10% level if* $\sqrt{T}|\hat{\rho}_1| > 1.65$. *With* $T = 100$, *we therefore need* $|\hat{\rho}_1| > 1.65/\sqrt{100} = 0.165$ *for rejection, and with* $T = 1000$ *we need* $|\hat{\rho}_1| > 1.65/\sqrt{1000} \approx 0.052$.

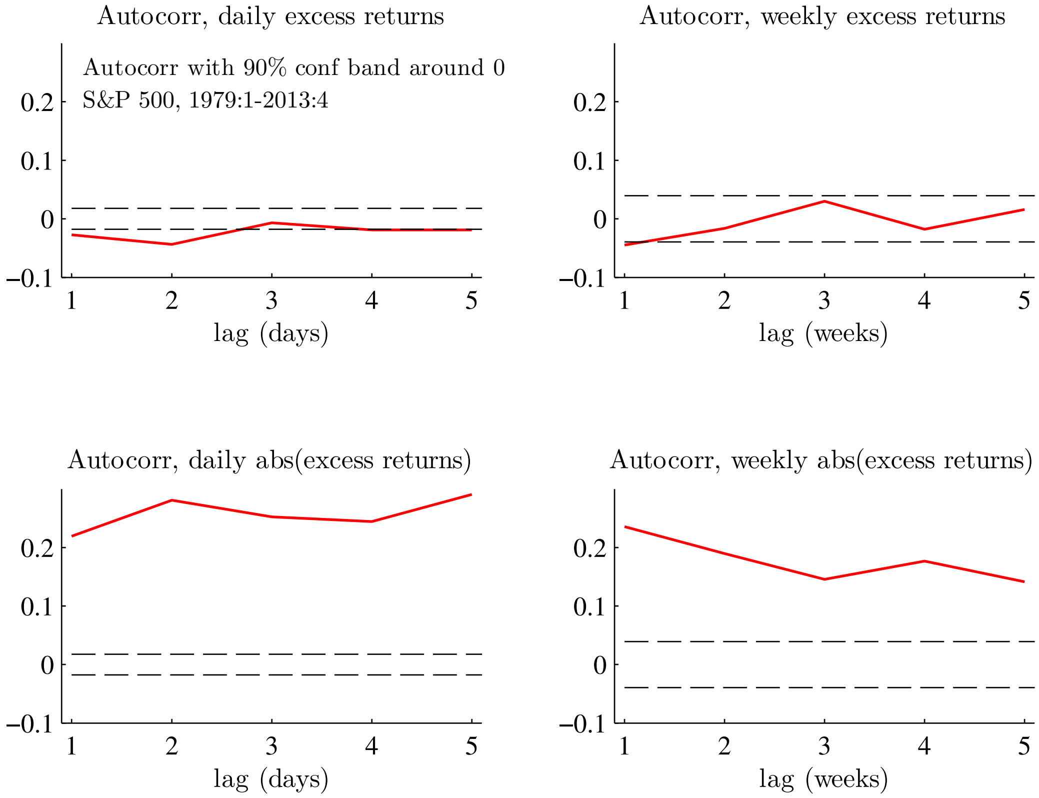


Figure 10.1: Predictability of US stock returns

10.2.2 Autoregressions

An alternative way of testing autocorrelations is to estimate an AR model

$$y_t = c + a_1 y_{t-1} + a_2 y_{t-2} + ... + a_p y_{t-p} + \varepsilon_t, \tag{10.10}$$

and then test if all slope coefficients $(a_1, a_2, ..., a_p)$ are zero with a χ^2 or F test. This approach is somewhat less general than testing if all autocorrelations are zero, but most stationary time series processes can be well approximated by an AR of relatively low order.

See Figure 10.3 for an illustration.

The autoregression can also allow for the coefficients to depend on the market situation. For instance, consider an AR(1), but where the autoregression coefficient may be

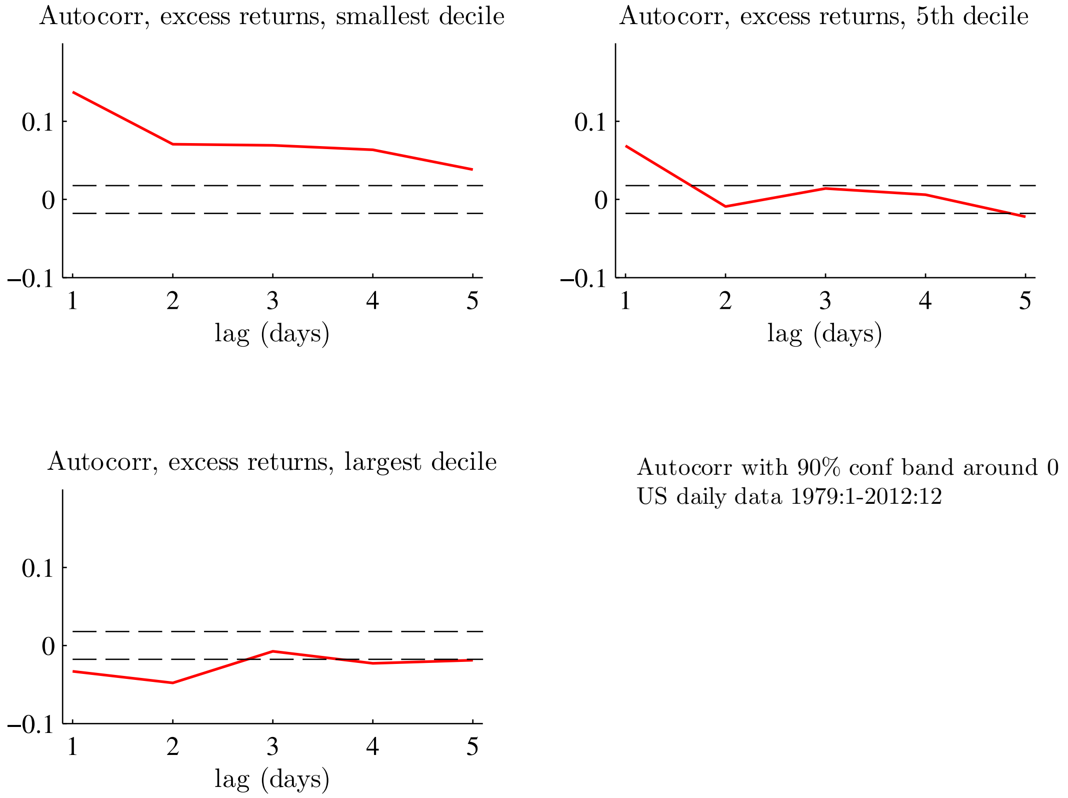


Figure 10.2: Predictability of US stock returns, size deciles

different depending on the sign of last period's return

$$y_t = c + a\delta(y_{t-1} \leq 0)y_{t-1} + b\delta(y_{t-1} > 0)y_{t-1} + \varepsilon_t, \text{ where} \tag{10.11}$$

$$\delta(q) = \begin{cases} 1 \text{ if } q \text{ is true} \\ 0 \text{ else.} \end{cases}$$

See Figure 10.4 for an illustration.

Inference of the slope coefficient in autoregressions on returns for longer data horizons than the data frequency (for instance, analysis of weekly returns in a data set consisting of daily observations) must be done with care. If only non-overlapping returns are used (use the weekly return for a particular weekday only, say Wednesdays), the standard LS expression for the standard deviation of the autoregressive parameter is likely to be reasonable. This is not the case, if overlapping returns (all daily data on weekly returns) are

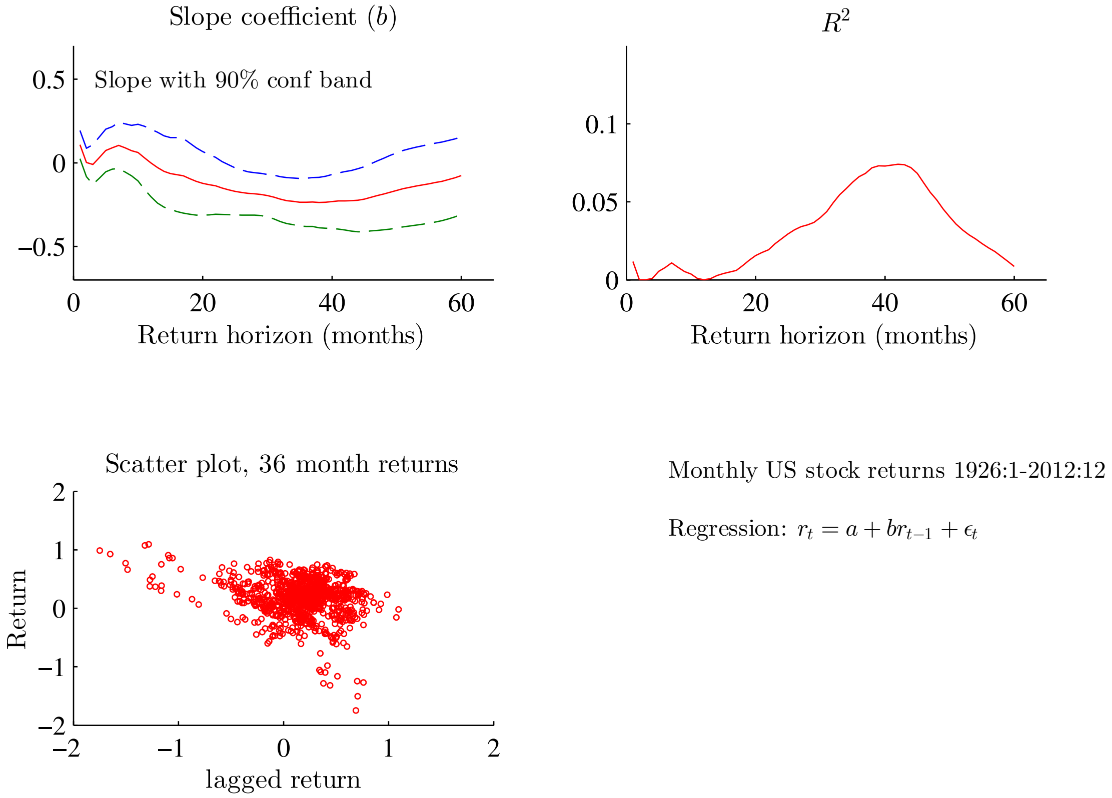


Figure 10.3: Predictability of US stock returns

used.

10.3 Other Predictors and Methods

There are many other possible predictors of future stock returns. For instance, both the dividend-price ratio and nominal interest rates have been used to predict long-run returns, and lagged short-run returns on other assets have been used to predict short-run returns.

10.3.1 Lead-Lags

Stock indices have more positive autocorrelation than (most) individual stocks: there should therefore be fairly strong cross-autocorrelations across individual stocks. Indeed, this is also what is found in US data where weekly returns of large size stocks forecast

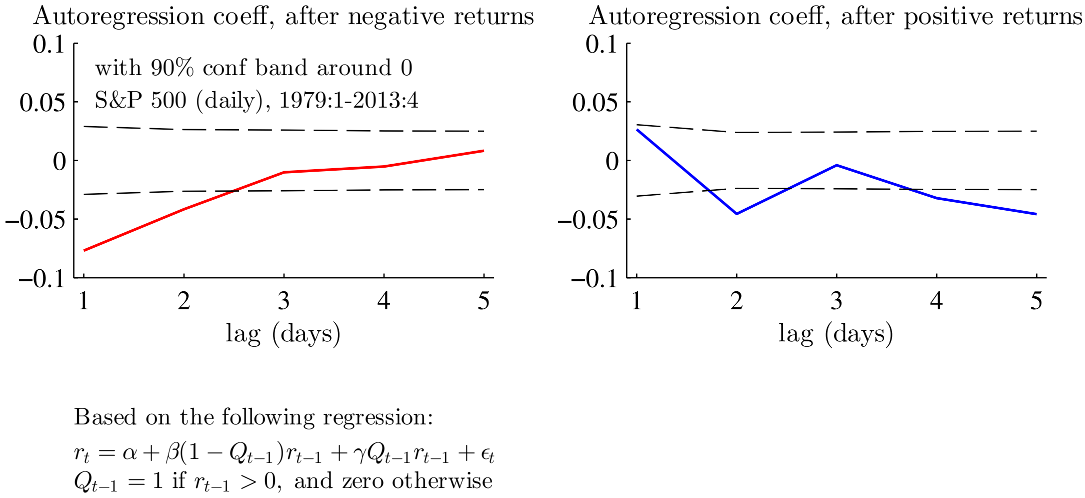


Figure 10.4: Predictability of US stock returns, results from a regression with interactive dummies

weekly returns of small size stocks. See Figure 10.5 for an illustration.

10.3.2 Dividend-Price Ratio as a Predictor

One of the most successful attempts to forecast long-run returns is a regression of future returns on the current dividend-price ratio (here in logs)

$$\sum_{s=1}^{q} r_{t+s} = \alpha + \beta_q(d_t - p_t) + \varepsilon_{t+q}. \tag{10.12}$$

See Figure 10.7 for an illustration.

10.4 Out-of-Sample Forecasting Performance

10.4.1 In-Sample versus Out-of-Sample Forecasting

To gauge the out-of-sample predictability, estimate the prediction equation using data for a moving data window up to and including $t-1$ (for instance, $t-W$ to $t-1$), and then make a forecast for period t. The forecasting performance of the equation is then compared with a benckmark model (eg. using the historical average as the predictor).

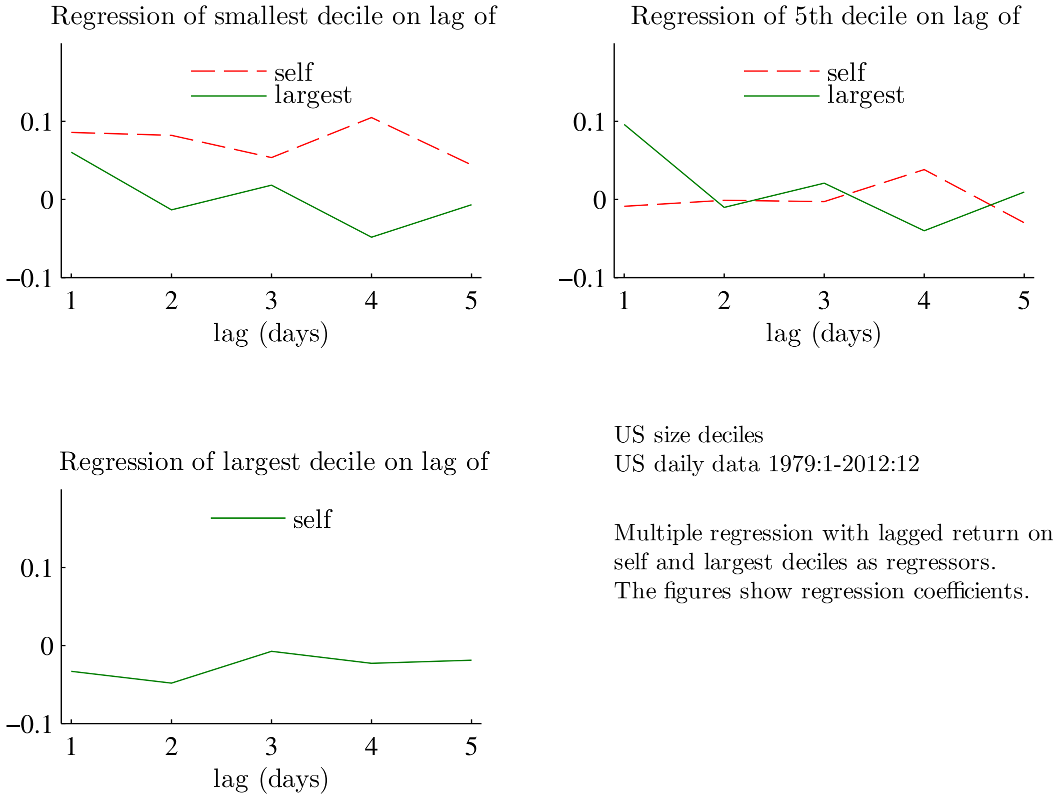


Figure 10.5: Coefficients from multiple prediction regressions

Notice that this benchmark model is also estimated on data up to an including $t-1$, so it changes over time.

To formalise the comparison, study the RMSE and the "out-of-sample R^2"

$$R_{OS}^2 = 1 - \frac{1}{T}\sum\nolimits_{t=s}^{T}(r_t - \hat{r}_t)^2 \Big/ \frac{1}{T}\sum\nolimits_{t=s}^{T}(r_t - \tilde{r}_t)^2, \tag{10.13}$$

where s is the first period with an out-of-sample forecast, $\hat{r}_t$ is the forecast based on the prediction model (estimated on data up to and including $t-1$) and $\tilde{r}_t$ is the prediction from some benchmark model (also estimated on data up to and including $t-1$).

Goyal and Welch (2008) find that the evidence of predictability of equity returns disappears when out-of-sample forecasts are considered.

See Figures 10.8 –10.10 for an illustration.

(Auto-)correlation matrix, monthly FF returns 1957:1-2012:12

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	0.18	0.19	0.18	0.15	0.15	0.22	0.20	0.20	0.17	0.15	0.23	0.21	0.18	0.20	0.16	0.22	0.20	0.18	0.18	0.16	0.18	0.18	0.16	0.14	0.15
2	0.17	0.17	0.18	0.16	0.16	0.21	0.19	0.20	0.18	0.16	0.22	0.22	0.19	0.22	0.18	0.21	0.21	0.19	0.19	0.18	0.17	0.18	0.17	0.15	0.17
3	0.17	0.18	0.19	0.17	0.17	0.20	0.20	0.20	0.19	0.17	0.21	0.21	0.20	0.22	0.18	0.21	0.21	0.20	0.20	0.18	0.18	0.18	0.18	0.17	0.17
4	0.18	0.20	0.20	0.19	0.19	0.21	0.21	0.22	0.21	0.19	0.22	0.23	0.22	0.23	0.20	0.23	0.23	0.22	0.22	0.20	0.19	0.20	0.19	0.18	0.18
5	0.22	0.23	0.25	0.23	0.24	0.25	0.25	0.26	0.26	0.25	0.26	0.27	0.26	0.28	0.25	0.26	0.27	0.27	0.27	0.25	0.22	0.24	0.23	0.24	0.25
6	0.12	0.12	0.12	0.10	0.10	0.15	0.14	0.14	0.12	0.10	0.18	0.16	0.14	0.16	0.11	0.18	0.16	0.14	0.14	0.13	0.15	0.14	0.12	0.11	0.12
7	0.13	0.14	0.15	0.13	0.13	0.16	0.15	0.16	0.15	0.13	0.18	0.18	0.16	0.18	0.15	0.18	0.18	0.17	0.16	0.15	0.15	0.15	0.13	0.14	0.15
8	0.12	0.13	0.14	0.12	0.12	0.15	0.15	0.16	0.15	0.13	0.17	0.18	0.17	0.19	0.16	0.18	0.19	0.18	0.18	0.15	0.16	0.16	0.15	0.16	0.16
9	0.12	0.13	0.14	0.13	0.13	0.14	0.14	0.15	0.15	0.14	0.15	0.18	0.17	0.19	0.16	0.17	0.18	0.18	0.18	0.16	0.14	0.15	0.15	0.16	0.17
10	0.12	0.14	0.15	0.14	0.15	0.15	0.16	0.17	0.17	0.16	0.16	0.18	0.19	0.20	0.19	0.17	0.19	0.20	0.19	0.19	0.16	0.17	0.17	0.18	0.20
11	0.07	0.07	0.08	0.06	0.06	0.11	0.09	0.10	0.08	0.06	0.13	0.13	0.10	0.13	0.08	0.13	0.13	0.11	0.11	0.10	0.12	0.11	0.10	0.10	0.10
12	0.10	0.11	0.13	0.10	0.10	0.13	0.13	0.14	0.12	0.11	0.15	0.16	0.14	0.16	0.13	0.16	0.17	0.16	0.15	0.12	0.15	0.13	0.12	0.13	0.13
13	0.11	0.12	0.14	0.12	0.12	0.14	0.13	0.15	0.14	0.12	0.15	0.16	0.15	0.17	0.14	0.17	0.17	0.16	0.16	0.13	0.14	0.12	0.12	0.13	0.14
14	0.08	0.09	0.12	0.10	0.11	0.12	0.12	0.13	0.13	0.12	0.13	0.15	0.14	0.16	0.14	0.15	0.16	0.17	0.15	0.14	0.14	0.13	0.13	0.15	0.15
15	0.10	0.11	0.13	0.11	0.11	0.12	0.13	0.13	0.13	0.12	0.13	0.15	0.15	0.16	0.14	0.15	0.15	0.16	0.15	0.13	0.13	0.12	0.11	0.13	0.15
16	0.08	0.08	0.08	0.05	0.05	0.10	0.09	0.09	0.07	0.05	0.12	0.11	0.08	0.10	0.05	0.12	0.10	0.09	0.08	0.07	0.10	0.09	0.07	0.06	0.07
17	0.09	0.11	0.12	0.10	0.10	0.13	0.13	0.13	0.13	0.10	0.14	0.15	0.13	0.15	0.12	0.15	0.15	0.15	0.13	0.11	0.12	0.11	0.11	0.11	0.11
18	0.09	0.10	0.12	0.10	0.10	0.11	0.12	0.12	0.12	0.10	0.12	0.14	0.12	0.14	0.11	0.14	0.14	0.15	0.13	0.11	0.11	0.10	0.10	0.11	0.11
19	0.07	0.08	0.11	0.09	0.09	0.10	0.10	0.11	0.11	0.10	0.11	0.13	0.11	0.14	0.11	0.13	0.13	0.14	0.12	0.11	0.12	0.10	0.11	0.11	0.13
20	0.08	0.09	0.12	0.10	0.12	0.10	0.12	0.13	0.13	0.12	0.11	0.14	0.14	0.15	0.13	0.13	0.14	0.15	0.14	0.13	0.12	0.11	0.10	0.12	0.14
21	0.05	0.06	0.07	0.05	0.04	0.08	0.08	0.08	0.06	0.04	0.11	0.09	0.06	0.08	0.03	0.11	0.07	0.07	0.05	0.04	0.09	0.06	0.04	0.03	0.03
22	0.05	0.06	0.08	0.06	0.05	0.08	0.08	0.08	0.08	0.06	0.09	0.10	0.08	0.10	0.06	0.10	0.09	0.09	0.06	0.06	0.07	0.07	0.05	0.06	0.06
23	0.03	0.04	0.06	0.04	0.04	0.05	0.06	0.06	0.06	0.05	0.06	0.08	0.07	0.08	0.04	0.08	0.07	0.07	0.05	0.05	0.05	0.05	0.05	0.05	0.05
24	0.04	0.04	0.06	0.05	0.05	0.06	0.06	0.06	0.07	0.06	0.07	0.08	0.07	0.08	0.06	0.08	0.08	0.09	0.06	0.06	0.07	0.05	0.05	0.05	0.07
25	0.07	0.07	0.09	0.07	0.08	0.10	0.10	0.10	0.10	0.09	0.11	0.12	0.11	0.11	0.11	0.13	0.12	0.11	0.10	0.09	0.13	0.09	0.09	0.09	0.09

Figure 10.6: Illustration of the cross-autocorrelations, $\text{Corr}(R_t, R_{t-k})$, monthly FF data. Dark colors indicate high correlations, light colors indicate low correlations.

10.4.2 Trading Strategies

Another way to measure predictability and to illustrate its economic importance is to calculate the return of a *dynamic trading strategy*, and then measure the "performance" of this strategy in relation to some benchmark portfolios. The trading strategy should, of course, be based on the variable that is supposed to forecast returns.

A common way (since Jensen, updated in Huberman and Kandel (1987)) is to study the performance of a portfolio by running the following regression

$$R_{1t} - R_{ft} = \alpha + \beta'(R_{mt} - R_{ft}) + \varepsilon_t, \text{ with} \tag{10.14}$$
$$\mathrm{E}\,\varepsilon_t = 0 \text{ and } \text{Cov}(R_{mt} - R_{ft}, \varepsilon_t) = 0,$$

where $R_{1t} - R_{ft}$ is the excess return on the portfolio being studied and $R_{mt} - R_{ft}$ the excess returns of a vector of benchmark portfolios (for instance, only the market portfolio if we want to rely on CAPM; returns times conditional information if we want to allow for time-variation in expected benchmark returns). Neutral performance (that is, that the tangency portfolio is unchanged and the two MV frontiers intersect there) requires $\alpha = 0$, which can be tested with a t test.

See Figure 10.11 for an illustration.

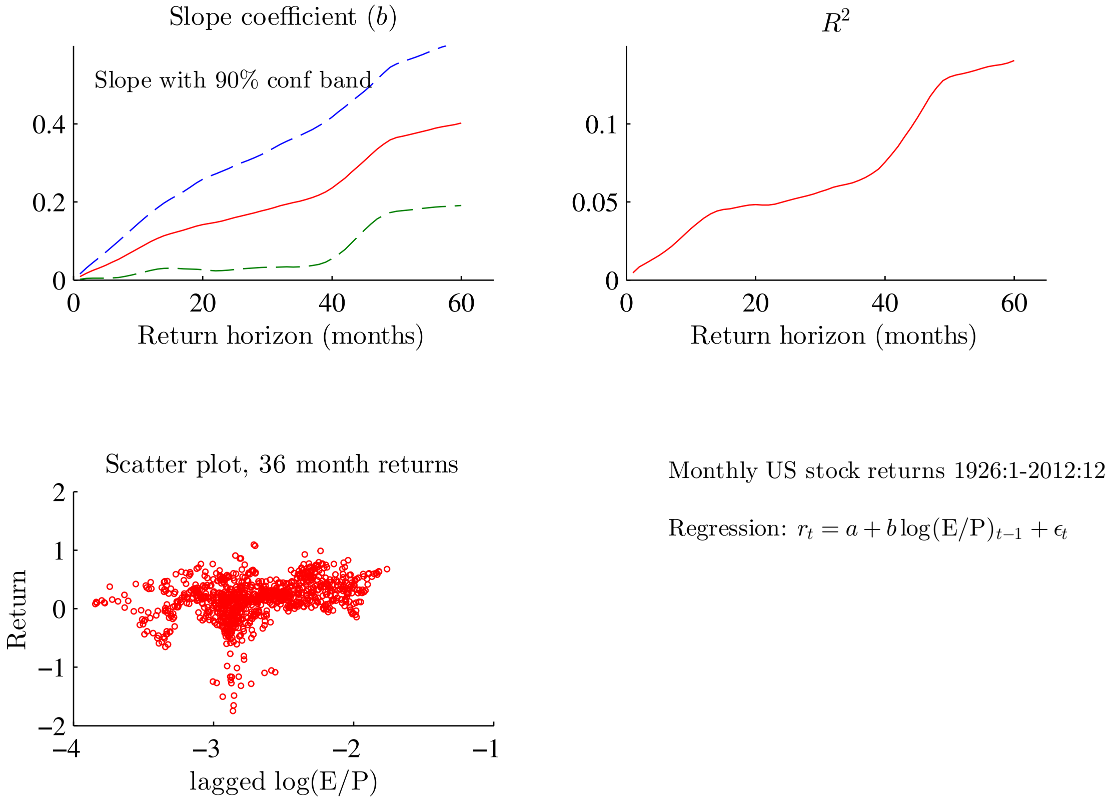


Figure 10.7: Predictability of US stock returns

10.4.3 Technical Analysis

Main reference: Bodie, Kane, and Marcus (2002) 12.2; Neely (1997) (overview, foreign exchange market)
Further reading: Murphy (1999) (practical, a believer's view); The Economist (1993) (overview, the perspective of the early 1990s); Brock, Lakonishok, and LeBaron (1992) (empirical, stock market); Lo, Mamaysky, and Wang (2000) (academic article on return distributions for "technical portfolios")

Technical analysis is typically a data mining exercise which looks for local trends or systematic non-linear patterns. The basic idea is that markets are not instantaneously efficient: prices react somewhat slowly and predictably to news. The logic is essentially that an observed price move must be due to some news (exactly which one is not very important) and that old patterns can tell us where the price will move in the near future.

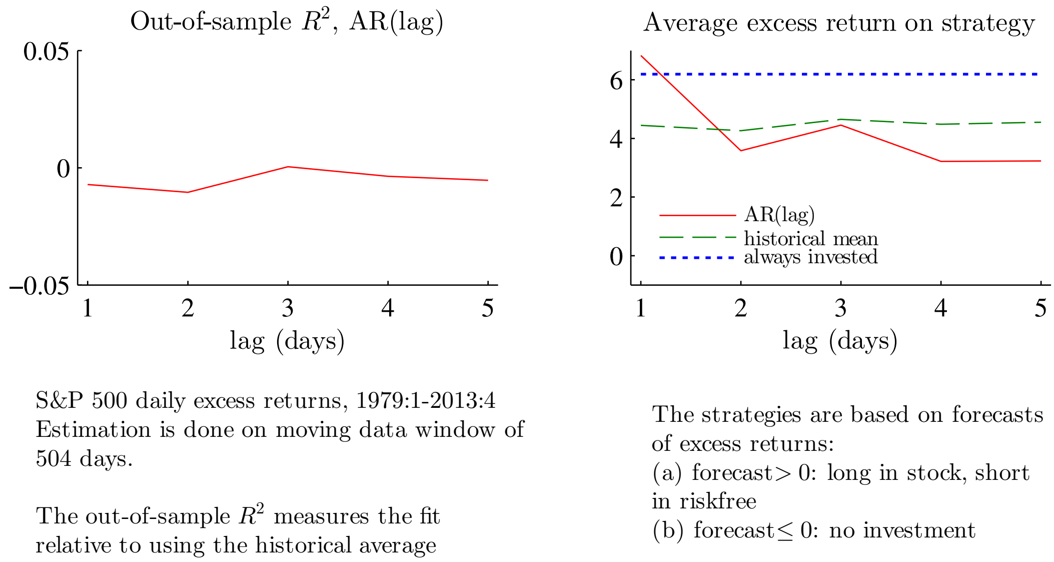


Figure 10.8: Short-run predictability of US stock returns, out-of-sample

This is an attempt to gather more detailed information than that used by the market as a whole. In practice, the technical analysis amounts to plotting different transformations (for instance, a moving average) of prices—and to spot known patterns. This section summarizes some simple trading rules that are used.

Many trading rules rely on some kind of local trend which can be thought of as positive autocorrelation in price movements (also called momentum[1]).

A *moving average rule* is to buy if a short moving average (equally weighted or exponentially weighted) goes above a long moving average. The idea is that event signals a new upward trend. Let S (L) be the lag order of a short (long) moving average, with $S < L$ and let b be a bandwidth (perhaps 0.01). Then, a MA rule for period t could be

$$\begin{bmatrix} \text{buy in } t \text{ if} & MA_{t-1}(S) > MA_{t-1}(L)(1+b) \\ \text{sell in } t \text{ if} & MA_{t-1}(S) < MA_{t-1}(L)(1-b) \\ \text{no change} & \text{otherwise} \end{bmatrix}, \text{ where} \tag{10.15}$$

$$MA_{t-1}(S) = (p_{t-1} + \ldots + p_{t-S})/S.$$

[1]In physics, momentum equals the mass times speed.

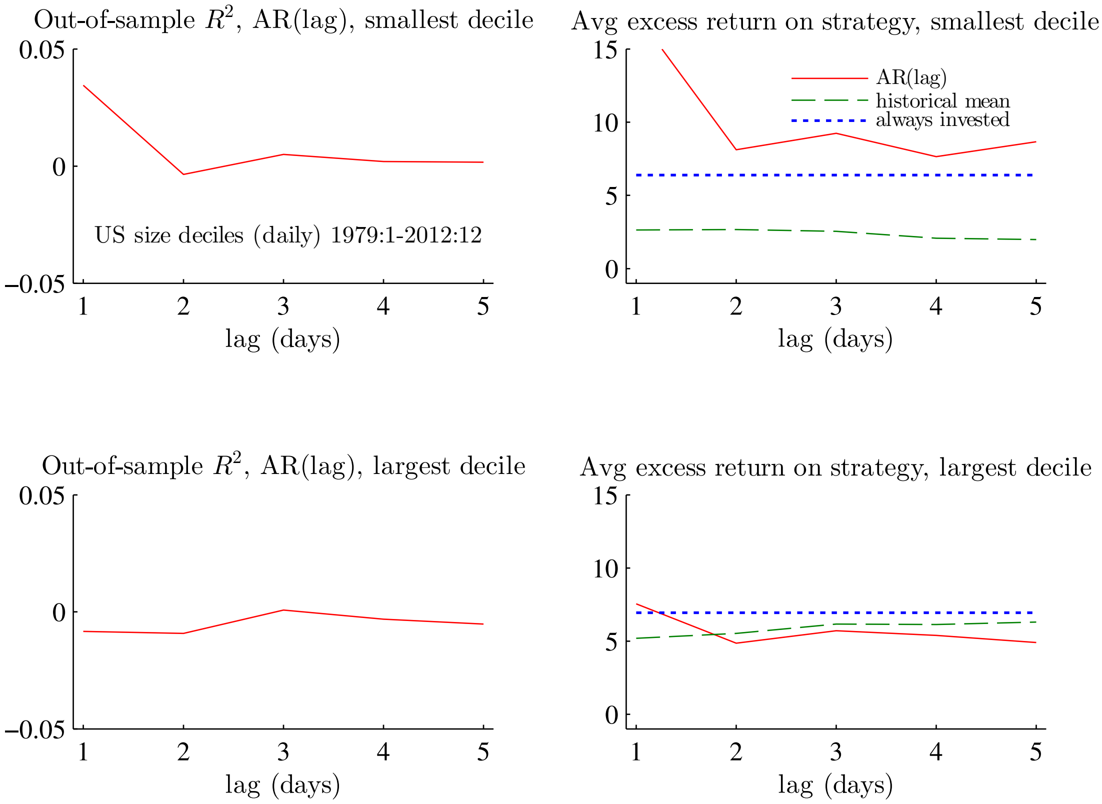


Figure 10.9: Short-run predictability of US stock returns, out-of-sample. See Figure 10.8 for details on the strategies.

The difference between the two moving averages is called an *oscillator*

$$\text{oscillator}_t = MA_t(S) - MA_t(L), \tag{10.16}$$

(or sometimes, moving average convergence divergence, MACD) and the sign is taken as a trading signal (this is the same as a moving average crossing, MAC).[2] A version of the moving average oscillator is the *relative strength index*[3], which is the ratio of average price level (or returns) on "up" days to the average price (or returns) on "down" days—during the last z (14 perhaps) days. Yet another version is to compare the oscillator$_t$ to an moving average of the oscillator (also called a signal line).

[2]Yes, the rumour is true: the tribe of chartists is on the verge of developing their very own language.

[3]Not to be confused with relative strength, which typically refers to the ratio of two different asset prices (for instance, an equity compared to the market).

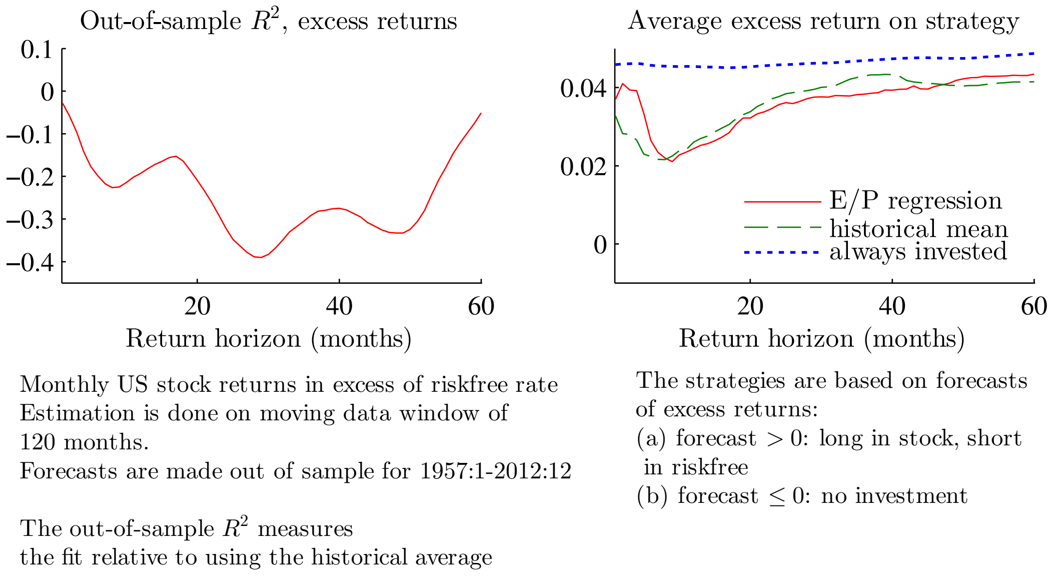


Figure 10.10: Long-run predictability of US stock returns, out-of-sample

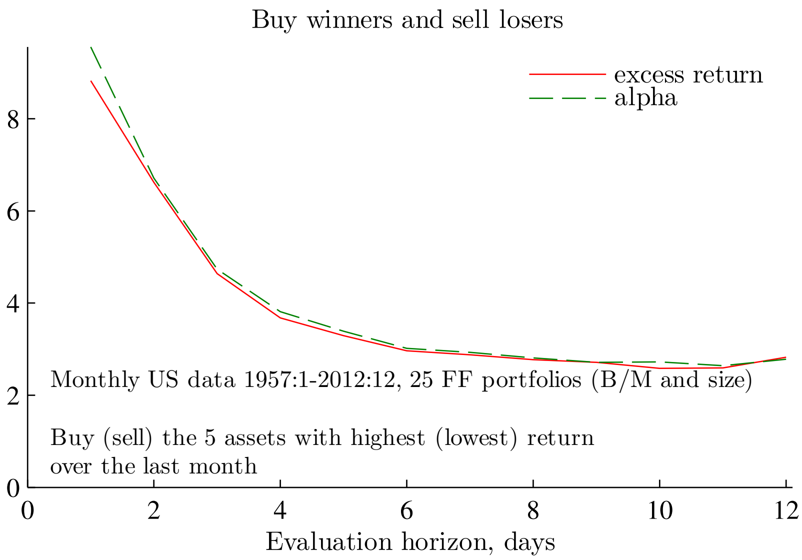


Figure 10.11: Predictability of US stock returns, momentum strategy

The *trading range break-out rule* typically amounts to buying when the price rises above a previous peak (local maximum). The idea is that a previous peak is a *resistance level* in the sense that some investors are willing to sell when the price reaches that value (perhaps because they believe that prices cannot pass this level; clear risk of circular reasoning or self-fulfilling prophecies; round numbers often play the role as resistance levels). Once this artificial resistance level has been broken, the price can possibly rise substantially. On the downside, a *support level* plays the same role: some investors are willing to buy when the price reaches that value. To implement this, it is common to let the resistance/support levels be proxied by minimum and maximum values over a data window of length L. With a bandwidth b (perhaps 0.01), the rule for period t could be

$$\begin{bmatrix} \text{buy in } t \text{ if} & P_t > M_{t-1}(1+b) \\ \text{sell in } t \text{ if} & P_t < m_{t-1}(1-b) \\ \text{no change} & \text{otherwise} \end{bmatrix}, \text{ where} \tag{10.17}$$

$$M_{t-1} = \max(p_{t-1}, \ldots, p_{t-S})$$
$$m_{t-1} = \min(p_{t-1}, \ldots, p_{t-S}).$$

When the price is already trending up, then the trading range break-out rule may be replaced by a *channel rule*, which works as follows. First, draw a *trend line* through previous lows and a *channel line* through previous peaks. Extend these lines. If the price moves above the channel (band) defined by these lines, then buy. A version of this is to define the channel by a *Bollinger band*, which is ± 2 standard deviations from a moving data window around a moving average.

If we instead believe in mean reversion of the prices, then we can essentially reverse the previous trading rules: we would typically sell when the price is high. See Figure 10.12 and Table 10.1.

	Mean	Std
All days	0.032	1.165
After buy signal	0.054	1.716
After neutral signal	0.047	0.943
After sell signal	0.007	0.903

Table 10.1: Returns (daily, in %) from technical trading rule (Inverted MA rule). Daily S&P 500 data 1990:1-2013:4

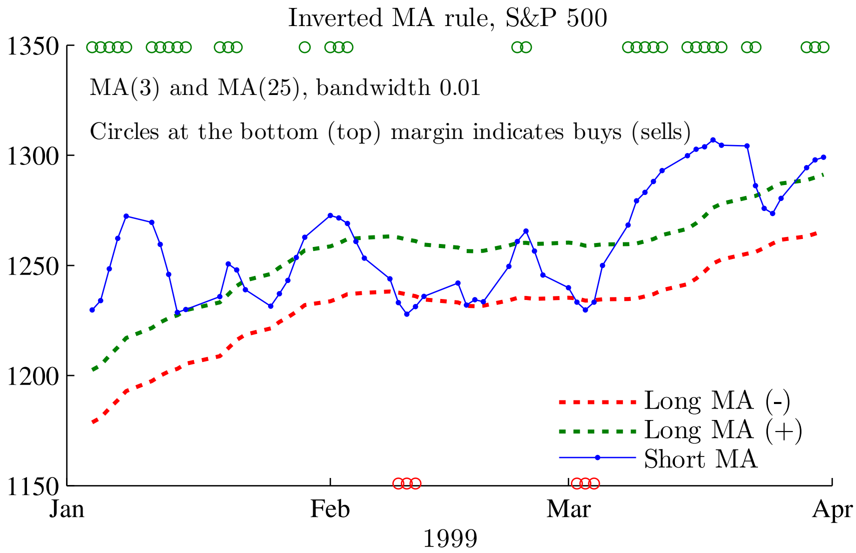


Figure 10.12: Examples of trading rules

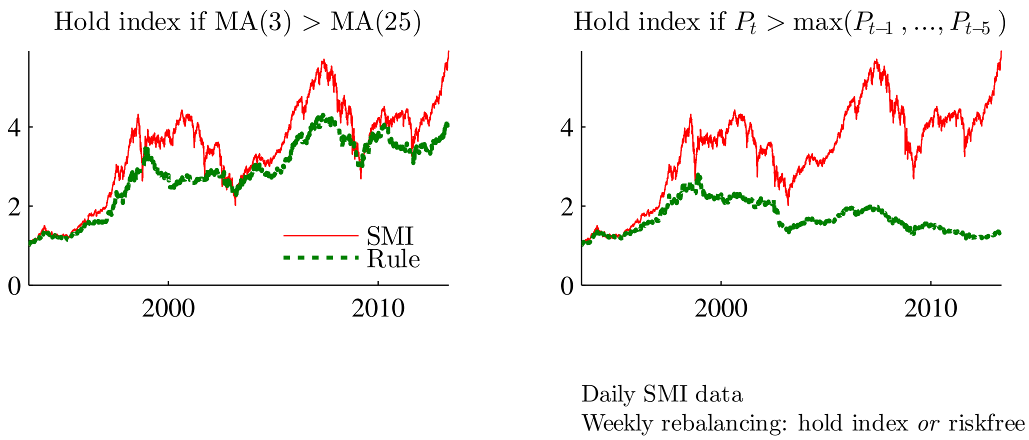


Figure 10.13: Examples of trading rules

10.5 Security Analysts

Reference: Makridakis, Wheelwright, and Hyndman (1998) 10.1 and Elton, Gruber, Brown, and Goetzmann (2010) 26

10.5.1 Evidence on Analysts' Performance

Makridakis, Wheelwright, and Hyndman (1998) 10.1 shows that there is little evidence that the average stock analyst beats (on average) the market (a passive index portfolio). In fact, less than half of the analysts beat the market. However, there are analysts which seem to outperform the market for some time, but the autocorrelation in over-performance is weak. The evidence from mutual funds is similar. For them it is typically also found that their portfolio weights do not anticipate price movements.

It should be remembered that many analysts also are sales persons: either of a stock (for instance, since the bank is underwriting an offering) or of trading services. It could well be that their objective function is quite different from minimizing the squared forecast errors—or whatever we typically use in order to evaluate their performance. (The number of litigations in the US after the technology boom/bust should serve as a strong reminder of this.)

10.5.2 Do Security Analysts Overreact?

The paper by Bondt and Thaler (1990) compares the (semi-annual) forecasts (one- and two-year time horizons) with actual changes in earnings per share (1976-1984) for several hundred companies. The paper has regressions like

$$\text{Actual change} = \alpha + \beta(\text{forecasted change}) + \text{ residual},$$

and then studies the estimates of the α and β coefficients. With rational expectations (and a long enough sample), we should have $\alpha = 0$ (no constant bias in forecasts) and $\beta = 1$ (proportionality, for instance no exaggeration).

The main findings are as follows. The main result is that $0 < \beta < 1$, so that the forecasted change tends to be too wild in a systematic way: a forecasted change of 1% is (on average) followed by a less than 1% actual change in the same direction. This means that analysts in this sample tended to be too extreme—to exaggerate both positive and negative news.

10.5.3 High-Frequency Trading Based on Recommendations from Stock Analysts

Barber, Lehavy, McNichols, and Trueman (2001) give a somewhat different picture. They focus on the profitability of a trading strategy based on analyst's recommendations.

They use a huge data set (some 360,000 recommendations, US stocks) for the period 1985-1996. They sort stocks in to five portfolios depending on the consensus (average) recommendation—and redo the sorting every day (if a new recommendation is published). They find that such a daily trading strategy gives an annual 4% abnormal return on the portfolio of the most highly recommended stocks, and an annual -5% abnormal return on the least favourably recommended stocks.

This strategy requires a lot of trading (a turnover of 400% annually), so trading costs would typically reduce the abnormal return on the best portfolio to almost zero. A less frequent rebalancing (weekly, monthly) gives a very small abnormal return for the best stocks, but still a negative abnormal return for the worst stocks. Chance and Hemler (2001) obtain similar results when studying the investment advise by 30 professional "market timers."

10.5.4 Economic Experts

Several papers, for instance, Bondt (1991) and Söderlind (2010), have studied whether economic experts can predict the broad stock markets. The results suggests that they cannot. For instance, Söderlind (2010) show that the economic experts that participate in the semi-annual Livingston survey (mostly bank economists) *(ii)* forecast the S&P worse than the historical average (recursively estimated), and that their forecasts are strongly correlated with recent market data (which in itself, cannot predict future returns).

10.5.5 Analysts and Industries

Boni and Womack (2006) study data on some 170,000 recommendations for a very large number of U.S. companies for the period 1996–2002. Focusing on revisions of recommendations, the papers shows that analysts are better at ranking firms within an industry than ranking industries.

10.5.6 Insiders

Corporate insiders *used to* earn superior returns, mostly driven by selling off stocks before negative returns. (There is little/no systematic evidence of insiders gaining by buying before high returns.) Actually, investors who followed the insider's registered transactions (in the U.S., these are made public six weeks after the reporting period), also used to earn some superior returns. It seems as if these patterns have more or less disappeared.

10.6 Event Studies

Reference: Bodie, Kane, and Marcus (2005) 12.3 or Copeland, Weston, and Shastri (2005) 11
Reference (advanced): Campbell, Lo, and MacKinlay (1997) 4

10.6.1 Basic Structure

The idea of an event study is to study the effect (on returns) of a special event by using a cross-section of such events. For instance, what is the effect of a negative earnings surprise on the share price?

According to the efficient market hypothesis, only news should move the asset price, so it is often necessary to explicitly model the previous expectations to define the event. For earnings, the event is typically taken to be a dummy that indicates if the earnings announcement is smaller than (some average of) analysts' forecast.

To isolate the effect of the event, we study the abnormal return of asset i in period t

$$u_{it} = R_{it} - R_{it}^{normal}, \tag{10.18}$$

where R_{it} is the actual return and the last term is the normal return (which may differ across assets and time). The definition of the normal return is discussed in detail in Section 10.6.2.

Suppose we have a sample of n such events. To keep the notation simple, we "normalize" the time so period 0 is the time of the event (irrespective of its actual calendar time).

To control for information leakage and slow price adjustment, the abnormal return is often calculated for some time before and after the event: the "event window" (often ± 20 days or so). For day s (that is, s days after the event time 0), the cross sectional average abnormal return is

$$\bar{u}_s = \textstyle\sum_{i=1}^{n} u_{is}/n. \tag{10.19}$$

For instance, $\bar{u}_2$ is the average abnormal return two days after the event, and $\bar{u}_{-1}$ is for one day before the event.

The cumulative abnormal return (CAR) of asset i is simply the sum of the abnormal return in (10.18) over some period around the event. It is often calculated from the beginning of the event window. For instance, if the event window starts at $-w$, then the q-period (day?) car for firm i is

$$\text{car}_{iq} = u_{i,-w} + u_{i,-w+1} + \ldots + u_{i,-w+q-1}. \tag{10.20}$$

The cross sectional average of the q-period car is

$$\overline{\text{car}}_q = \textstyle\sum_{i=1}^{n} \text{car}_{iq}/n. \tag{10.21}$$

See Figure 10.14 for an empirical example.

Example 10.4 *(Abnormal returns for $\pm$ day around event, two firms) Suppose there are two firms and the event window contains ± 1 day around the event day, and that the abnormal returns (in percent) are*

Time	*Firm 1*	*Firm 2*	*Cross-sectional Average*
−1	0.2	−0.1	0.05
0	1.0	2.0	1.5
1	0.1	0.3	0.2

We have the following cumulative returns

Time	*Firm 1*	*Firm 2*	*Cross-sectional Average*
−1	0.2	−0.1	0.05
0	1.2	1.9	1.55
1	1.3	2.2	1.75

10.6.2 Models of Normal Returns

This section summarizes the most common ways of calculating the normal return in (10.18). The parameters in these models are typically estimated on a recent sample, the "estimation window," which ends before the event window. See Figure 10.15 for an illustration. In this way, the estimated behaviour of the normal return should be unaffected by the event. It is almost always assumed that the event is exogenous in the sense that it

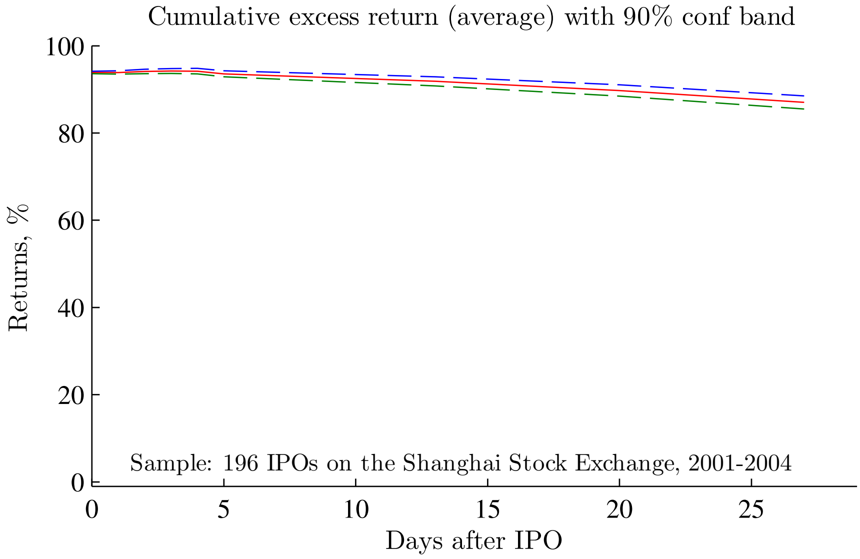


Figure 10.14: Event study of IPOs in Shanghai 2001–2004. (Data from Nou Lai.)

is not due to the movements in the asset price during either the estimation window or the event window.

The *constant mean return model* assumes that the return of asset i fluctuates randomly around some mean μ_i

$$\begin{aligned} R_{it} &= \mu_i + \varepsilon_{it} \text{ with} \\ \mathrm{E}\,\varepsilon_{it} &= \mathrm{Cov}(\varepsilon_{it}, \varepsilon_{i,t-s}) = 0. \end{aligned} \tag{10.22}$$

This mean is estimated by the sample average (during the estimation window). The normal return in (10.18) is then the estimated mean. $\hat{\mu}_i$ so the abnormal return (in the estimation window) becomes $\hat{\varepsilon}_{it}$. During the event window, we calculate the abnormal return as

$$u_{it} = R_{it} - \hat{\mu}_i. \tag{10.23}$$

The standard error of this is estimated by the standard error of $\hat{\varepsilon}_{it}$ (in the estimation window).

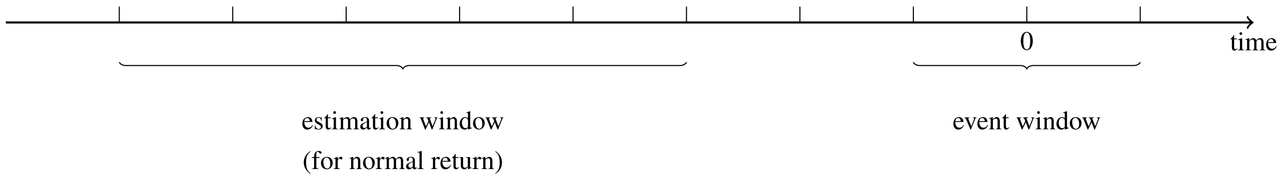


Figure 10.15: Event and estimation windows

The *market model* is a linear regression of the return of asset i on the market return

$$R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it} \text{ with} \tag{10.24}$$
$$\mathrm{E}\,\varepsilon_{it} = \mathrm{Cov}(\varepsilon_{it}, \varepsilon_{i,t-s}) = \mathrm{Cov}(\varepsilon_{it}, R_{mt}) = 0.$$

Notice that we typically do not impose the CAPM restrictions on the intercept in (10.24). The normal return in (10.18) is then calculated by combining the regression coefficients with the actual market return as $\hat{\alpha}_i + \hat{\beta}_i R_{mt}$, so the the abnormal return in the estimation window is $\hat{\varepsilon}_{it}$. For the event window we calculate the abnormal return as

$$u_{it} = R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{mt}. \tag{10.25}$$

The standard error of this is estimated by the standard error of $\hat{\varepsilon}_{it}$ (in the estimation window).

When we restrict $\alpha_i = 0$ and $\beta_i = 1$, then this approach is called the *market-adjusted-return model.* This is a particularly useful approach when there is no return data before the event, for instance, with an IPO. For the event window we calculate the abnormal return as

$$u_{it} = R_{it} - R_{mt} \tag{10.26}$$

and the standard error of it is estimated by $\mathrm{Std}(R_{it} - R_{mt})$ in the estimation window. This approach is especially convenient if there is no data in the estimation window (for instance, there is no return data before an IPO).

Yet another approach is to construct a normal return as the actual return on assets which are very similar to the asset with an event. For instance, if asset i is a small manufacturing firm (with an event), then the normal return could be calculated as the actual return for other small manufacturing firms (without events). In this case, the abnormal return becomes the difference between the actual return and the return on the matching portfolio. This type of *matching portfolio* is becoming increasingly popular. For the event window we calculate the abnormal return as

$$u_{it} = R_{it} - R_{pt}, \tag{10.27}$$

where R_{pt} is the return of the matching portfolio. The standard error of it is estimated by $\text{Std}(R_{it} - R_{pt})$ in the estimation window.

High frequency data can be very helpful, provided the time of the event is known. High frequency data effectively allows us to decrease the volatility of the abnormal return since it filters out irrelevant (for the event study) shocks to the return while still capturing the effect of the event.

10.6.3 Testing the Abnormal Return

It is typically assumed that the abnormal returns are uncorrelated across time and across assets. The first assumption is motivated by the very low autocorrelation of returns. The second assumption makes a lot of sense if the events are not overlapping in time, so that the event of assets i and j happen at different (calendar) times.

Let $\sigma_i^2 = \text{Var}(u_{it})$ be the variance of the abnormal return of asset i. The *variance of the cross-sectional* (across the n assets) *average*, $\bar{u}_s$ in (10.19), is then

$$\text{Var}(\bar{u}_s) = \textstyle\sum_{i=1}^{n} \sigma_i^2 / n^2, \tag{10.28}$$

since all covariances are assumed to be zero. In a large sample, we can therefore use a t-test since

$$\bar{u}_s / \text{Std}(\bar{u}_s) \to^d N(0, 1). \tag{10.29}$$

The *cumulative abnormal return* over q period, $\text{car}_{i,q}$, can also be tested with a t-test. Since the returns are assumed to have no autocorrelation the variance of the $\text{car}_{i,q}$

$$\text{Var}(\text{car}_{iq}) = q\sigma_i^2. \tag{10.30}$$

This variance is increasing in q since we are considering cumulative returns (not the time average of returns).

The *cross-sectional average* $car_{i,q}$ is then (similarly to (10.28))

$$\text{Var}(\overline{\text{car}}_q) = q\sum_{i=1}^{n}\sigma_i^2/n^2, \tag{10.31}$$

if the abnormal returns are uncorrelated across time and assets.

Example 10.5 *(Variances of abnormal returns) If the standard deviations of the daily abnormal returns of the two firms in Example 10.4 are* $\sigma_1 = 0.1$ *and and* $\sigma_2 = 0.2$*, then we have the following variances for the abnormal returns at different days*

Time	*Firm 1*	*Firm 2*	*Cross-sectional Average*
-1	0.1^2	0.2^2	$\left(0.1^2+0.2^2\right)/4$
0	0.1^2	0.2^2	$\left(0.1^2+0.2^2\right)/4$
1	0.1^2	0.2^2	$\left(0.1^2+0.2^2\right)/4$

Similarly, the variances for the cumulative abnormal returns are

Time	*Firm 1*	*Firm 2*	*Cross-sectional Average*
-1	0.1^2	0.2^2	$\left(0.1^2+0.2^2\right)/4$
0	2×0.1^2	2×0.2^2	$2\times\left(0.1^2+0.2^2\right)/4$
1	3×0.1^2	3×0.2^2	$3\times\left(0.1^2+0.2^2\right)/4$

Example 10.6 *(Tests of abnormal returns) By dividing the numbers in Example 10.4 by the square root of the numbers in Example 10.5 (that is, the standard deviations) we get the test statistics for the abnormal returns*

Time	*Firm 1*	*Firm 2*	*Cross-sectional Average*
-1	2	-0.5	0.4
0	10	10	13.4
1	1	1.5	1.8

Similarly, the variances for the cumulative abnormal returns we have

Time	*Firm 1*	*Firm 2*	*Cross-sectional Average*
-1	2	-0.5	0.4
0	8.5	6.7	9.8
1	7.5	6.4	9.0

Bibliography

Ang, J. S., and S. J. Ciccone, 2001, "International differences in analyst forecast properties," mimeo, Florida State University.

Barber, B., R. Lehavy, M. McNichols, and B. Trueman, 2001, "Can investors profit from the prophets? Security analyst recommendations and stock returns," *Journal of Finance*, 56, 531–563.

Bodie, Z., A. Kane, and A. J. Marcus, 2002, *Investments*, McGraw-Hill/Irwin, Boston, 5th edn.

Bodie, Z., A. Kane, and A. J. Marcus, 2005, *Investments*, McGraw-Hill, Boston, 6th edn.

Bolliger, G., 2001, "The characteristics of individual analysts' forecasts in Europe," mimeo, University of Neuchatel.

Bondt, W. F. M. D., 1991, "What do economists know about the stock market?," *Journal of Portfolio Management*, 17, 84–91.

Bondt, W. F. M. D., and R. H. Thaler, 1990, "Do security analysts overreact?," *American Economic Review*, 80, 52–57.

Boni, L., and K. L. Womack, 2006, "Analysts, industries, and price momentum," *Journal of Financial and Quantitative Analysis*, 41, 85–109.

Brock, W., J. Lakonishok, and B. LeBaron, 1992, "Simple technical trading rules and the stochastic properties of stock returns," *Journal of Finance*, 47, 1731–1764.

Brockwell, P. J., and R. A. Davis, 1991, *Time series: theory and methods*, Springer Verlag, New York, second edn.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, 1997, *The econometrics of financial markets*, Princeton University Press, Princeton, New Jersey.

Campbell, J. Y., and S. B. Thompson, 2008, "Predicting the equity premium out of sample: can anything beat the historical average," *Review of Financial Studies*, 21, 1509–1531.

Chance, D. M., and M. L. Hemler, 2001, "The performance of professional market timers: daily evidence from executed strategies," *Journal of Financial Economics*, 62, 377–411.

Cochrane, J. H., 2001, *Asset pricing*, Princeton University Press, Princeton, New Jersey.

Copeland, T. E., J. F. Weston, and K. Shastri, 2005, *Financial theory and corporate policy*, Pearson Education, 4 edn.

Cuthbertson, K., 1996, *Quantitative financial economics*, Wiley, Chichester, England.

Ederington, L. H., and J. C. Goh, 1998, "Bond rating agencies and stock analysts: who knows what when?," *Journal of Financial and Quantitative Analysis*, 33, 569–585.

Elton, E. J., M. J. Gruber, S. J. Brown, and W. N. Goetzmann, 2010, *Modern portfolio theory and investment analysis*, John Wiley and Sons, 8th edn.

Fama, E. F., and K. R. French, 1993, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3–56.

Goyal, A., and I. Welch, 2008, "A comprehensive look at the empirical performance of equity premium prediction," *Review of Financial Studies 2008*, 21, 1455–1508.

Huberman, G., and S. Kandel, 1987, "Mean-variance spanning," *Journal of Finance*, 42, 873–888.

Lo, A. W., H. Mamaysky, and J. Wang, 2000, "Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation," *Journal of Finance*, 55, 1705–1765.

Makridakis, S., S. C. Wheelwright, and R. J. Hyndman, 1998, *Forecasting: methods and applications*, Wiley, New York, 3rd edn.

Murphy, J. J., 1999, *Technical analysis of the financial markets*, New York Institute of Finance.

Neely, C. J., 1997, "Technical analysis in the foreign exchange market: a layman's guide," *Federal Reserve Bank of St. Louis Review*.

Priestley, M. B., 1981, *Spectral analysis and time series*, Academic Press.

Söderlind, P., 2010, “Predicting stock price movements: regressions versus economists,” *Applied Economics Letters*, 17, 869–874.

The Economist, 1993, “Frontiers of finance,” pp. 5–20.

11 Dynamic Portfolio Choice

More advanced material is denoted by a star (*). It is not required reading.

11.1 Optimal Portfolio Choice: CRRA Utility and iid Returns

Suppose the investor wants choose portfolio weights (v_t) to maximize expected utility, that is, to solve

$$\max_{v_t} \mathrm{E}_t\, u(W_{t+q}), \tag{11.1}$$

where and E_t denotes the expectations formed today, $u()$ is a utility function and W_{t+q} is the wealth (in real terms) at time $t + q$.

This is a standard (static) problem if the investor cannot (or it is too costly to) rebalance the portfolio. (In some cases this leads to a mean-variance portfolio, in other cases not.) If the distribution of assets returns is iid, then the portfolio choice is unchanged over time—otherwise it changes. For instance, with mean-variance preferences, the tangency portfolio changes as the expected returns and/or the covariance matrix do.

Instead, if the investor can rebalance the portfolio in every time period ($t + 1, ..., t + q - 1$), then this is a truly dynamic problem—which is typically more difficult to solve. However, when the utility function has constant relative risk aversion (CRRA) and returns are iid, then we know that the optimal portfolio weights are constant across time and independent of the investment horizon (q). We can then solve this as a standard static problem. The intuition for this result is straightforward: CRRA utility implies that the portfolio weights are independent of the wealth of the investor and iid returns imply that the outlook from today is the same as the outlook from yesterday, except that the investor might have gotten richer or poorer. (The same result holds if the objective function instead is to maximize the utility from stream of consumption, but with a CRRA utility function.)

With non-iid returns (predictability or time-varying volatility), the optimization is typically much more complicated. The next few sections present a few cases that we can handle.

11.2 Optimal Portfolio Choice: Logarithmic Utility and Non-iid Returns

Reference: Campbell and Viceira (2002)

11.2.1 The Optimization Problem 1

Let the objective in period t be to maximize the expected log wealth in some future period

$$\max \mathrm{E}_t \ln W_{t+q} = \max(\ln W_t + \mathrm{E}_t\, r_{t+1} + \mathrm{E}_t\, r_{t+2} + \ldots + \mathrm{E}_t\, r_{t+q}), \tag{11.2}$$

where r_t is the log return, $r_t = \ln(1 + R_t)$ where R_t is a net return. The investor can rebalance the portfolio weights every period.

Since the returns in the different periods enter separably, the best an investor can do in period t is to choose a portfolio that solves

$$\max \mathrm{E}_t\, r_{t+1}. \tag{11.3}$$

That is, to choose the one-period growth-optimal portfolio. But, a short run investor who maximizes $\mathrm{E}_t \ln[W_t(1 + R_{t+1})]) = \max(\ln W_t + \mathrm{E}_t\, r_{t+1})$ will choose the same portfolio, so there is no horizon effect. However, the portfolio choice may change over time, if the distribution of the returns do. (The same result holds if the objective function instead is to maximize the utility from stream of consumption, but with a logarithmic utility function.)

11.2.2 Approximating the Log Portfolio Return

In dynamic portfolio choice models it is often more convenient to work with logarithmic portfolio returns (since they are additive across time). This has a drawback, however, on the portfolio formation stage: the logarithmic portfolio return is not a linear function of the logarithmic returns of the assets in the portfolio. Therefore, we will use an approximation (which gets more and more precise as the length of the time interval decreases).

If there is only one risky asset and one riskfree asset, then $R_{pt} = vR_t + (1 - v)R_{ft}$. Let $r_{it} = \ln(1 + R_{it})$ denote the log return. Campbell and Viceira (2002) approximate the log portfolio return by

$$r_{pt} \approx r_{ft} + v\left(r_t - r_{ft}\right) + v\sigma^2/2 - v^2\sigma^2/2, \tag{11.4}$$

where σ^2 is the conditional variance of r_t. (That is, σ^2 is the variance of u_t in $r_t = \mathrm{E}_{t-1} r_t + u_t$.) Instead, if we let r_t denote an $n \times 1$ vector of risky log returns and v the portfolio weights, then the multivariate version is

$$r_{pt} \approx r_{ft} + v' \left(r_t - r_{ft} \right) + v'\sigma^2/2 - v'\Sigma v/2, \tag{11.5}$$

where Σ is the $n \times n$ covariance matrix of r_t and σ^2 is the $n \times 1$ vector of the variances (that is, the the diagonal elements of that covariance matrix). The portfolio weights, variances and covariances could be time-varying (and should then perhaps carry time subscripts).

Proof. (of (11.4)*) The portfolio return $R_p = vR_1 + (1 - v)R_f$ can be used to write

$$\frac{1 + R_p}{1 + R_f} = 1 + v \left(\frac{1 + R_1}{1 + R_f} - 1 \right).$$

The logarithm is

$$r_p - r_f = \ln \left\{ 1 + v \left[\exp(r_1 - r_f) - 1 \right] \right\}.$$

The function $f(x) = \ln \{1 + v [\exp(x) - 1]\}$ has the following derivatives (evaluated at $x = 0$): $df(x)/dx = v$ and $d^2 f(x)/dx^2 = v(1 - v)$, and notice that $f(0) = 0$. A second order Taylor approximation of the log portfolio return around $r_1 - r_f = 0$ is then

$$r_p - r_f = v \left(r_1 - r_f \right) + \frac{1}{2} v(1 - v) \left(r_1 - r_f \right)^2.$$

In a continuous time model, the square would equal its expectation, $\mathrm{Var}(r_1)$, so this further approximation is used to give (11.4). (The proof of (11.5) is just a multivariate extension of this.) ■

11.2.3 The Optimization Problem 2

The objective is to maximize the (conditional) expected value of the portfolio return as in (11.3). When there is one risky asset and a riskfree asset, then the portfolio return is given by the approximation (11.4). To simplify the notation a bit, let μ^e_{t+1} be the conditional expected excess return $\mathrm{E}_t(r_{t+1} - r_{f,t+1})$ and let σ^2_{t+1} be the conditional variance $(\mathrm{Var}_t(r_{t+1}))$. Notice that these moments are conditional on the information in t (when the portfolio decision is made) but refer to the returns in $t + 1$.

The optimization problem is then

$$\max_{v_t} r_{f,t+1} + v_t \mu^e_{t+1} + v_t \sigma^2_{t+1}/2 - v_t^2 \sigma^2_{t+1}/2. \tag{11.6}$$

The first order condition is

$$\begin{aligned} 0 &= \mu^e_{t+1} + \sigma^2_{t+1}/2 - v_t \sigma^2_{t+1}, \text{ so} \\ v_t &= \frac{\mu^e_{t+1} + \sigma^2_{t+1}/2}{\sigma^2_{t+1}}, \end{aligned} \tag{11.7}$$

which is very similar to a mean-variance portfolio choice. Clearly, the weight on the risky asset will change over time—if the expected excess return and/or the volatility does. We could think of the portfolio with v_t of the risky asset and $1 - v_t$ of the riskfree asset as a *managed portfolio*.

Example 11.1 *(Portfolio weight, single risky asset) Suppose* $\mu^e_{t+1} = 0.05$ *and* $\sigma^2_{t+1} = 0.15$*, then we have* $v_t = (0.05 + 0.15/2)/0.15 = 5/6 \approx 0.83$.

With many risky assets, the optimization problem is to maximize the expected value of (11.5). The optimal $n \times 1$ vector of portfolio weights is then

$$v_t = \Sigma^{-1}_{t+1}(\mu^e_{t+1} + \sigma^2_{t+1}/2), \tag{11.8}$$

where Σ_{t+1} is the conditional covariance matrix ($\text{Cov}_t(r_{t+1})$) and σ^2_{t+1} the $n \times 1$ vector of conditional variances. The weight on the riskfree asset is the remainder ($1 - \mathbf{1}' v_t$, where $\mathbf{1}$ is a vector of ones).

Proposition 11.2 *If the log returns are normally distributed, then (11.8) gives a portfolio on the mean-variance frontier of returns (not of log returns).*

Figures 11.1–11.2 illustrate mean returns and standard deviations, estimated by exponentially moving averages (as by RiskMetrics). Figures 11.3–11.4 show how the optimal portfolio weights change (assuming mean-variance preferences). It is clear that the portfolio weights change very dramatically—perhaps too much to be realistic. The portfolio weights seem to be particularly sensitive to movements in the average returns, which potentially a problem since the averages are often considered to be more difficult to estimate (with good precision) than the covariance matrix.

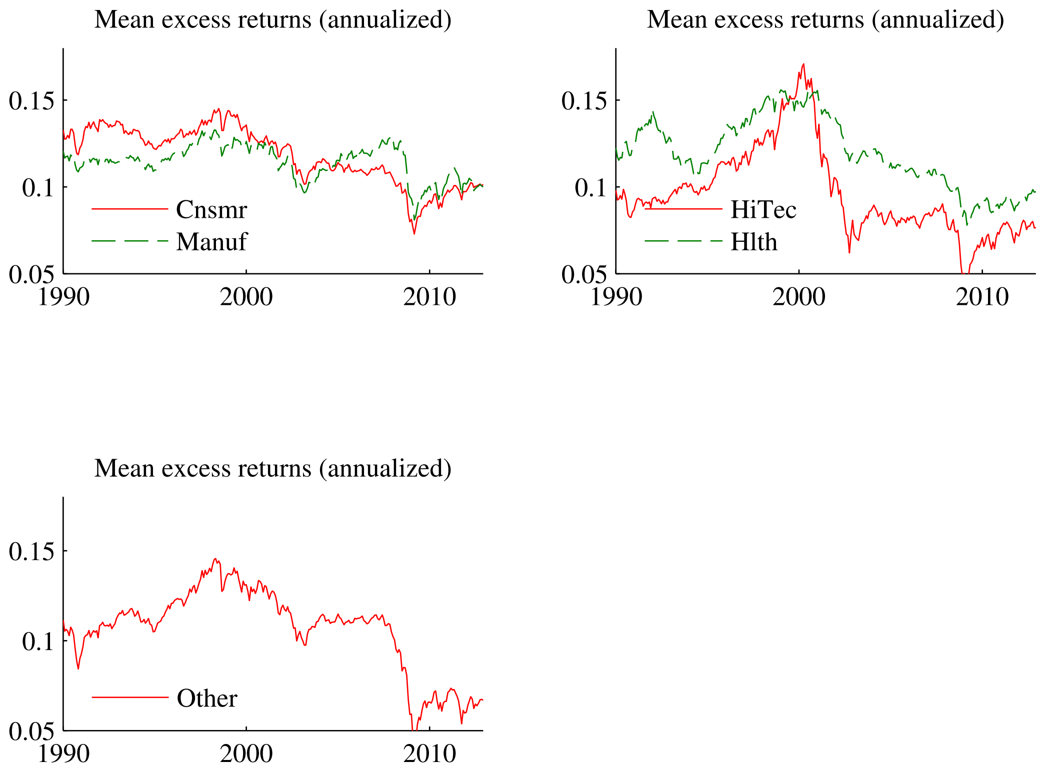


Figure 11.1: Dynamically updated estimates, 5 U.S. industries

Proof. (of (11.8)) From (11.5) we have

$$\mathrm{E}\, r_p \approx r_f + v'\mu^e + v'\sigma^2/2 - v'\Sigma v/2,$$

so the first order conditions are

$$\mu^e + \sigma^2/2 - \Sigma^{-1}v = \mathbf{0}_{n\times 1}.$$

Solve for v. ■

Proof. (of Proposition 11.2) First, notice that if the log return r_t in (11.5) is normally distributed, then so is the log portfolio return (r_{pt}). Second, recall that if $\ln y \sim N(\mu, \sigma^2)$, then $\mathrm{E}\, y = \exp\left(\mu + \sigma^2/2\right)$ and $\mathrm{Std}\,(y)\,/\,\mathrm{E}\, y = \sqrt{\exp(\sigma^2) - 1}$, so that $\ln \mathrm{E}\, y - \sigma^2/2 =$

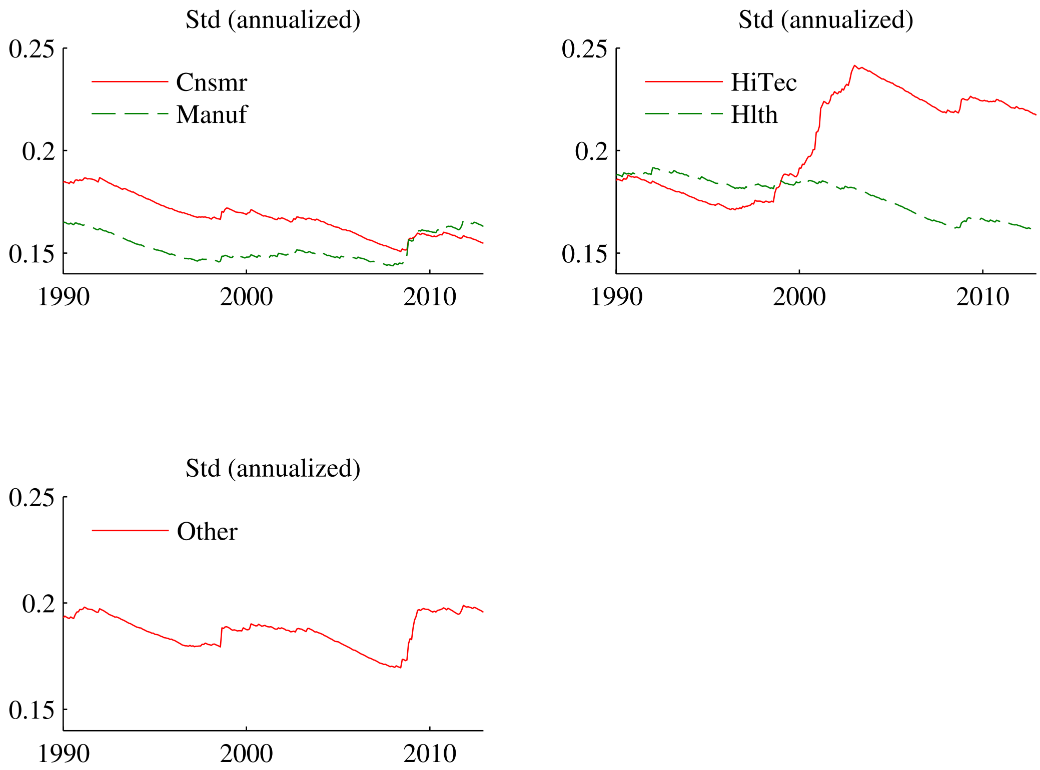


Figure 11.2: Dynamically updated estimates, 5 U.S. industries

μ and $\ln[\mathrm{Var}\,(y)\,/(\mathrm{E}\,y)^2 + 1] = \sigma^2$. Combine to write

$$\mu = \ln \mathrm{E}\,y - \ln[\mathrm{Var}\,(y)\,/(\mathrm{E}\,y)^2 + 1]/2,$$

which is increasing in $\mathrm{E}\,y$ and decreasing in $\mathrm{Var}(y)$. To prove the statement, notice that y corresponds to the gross return and $\ln y$ to the log return, so μ corresponds to $\mathrm{E}_t\, r_{pt+1}$. Clearly, μ is increasing in $\mathrm{E}\,y$ and decreasing in $\mathrm{Var}(y)$, so the solution will be on the MV frontier of the (gross and net) portfolio return. ■

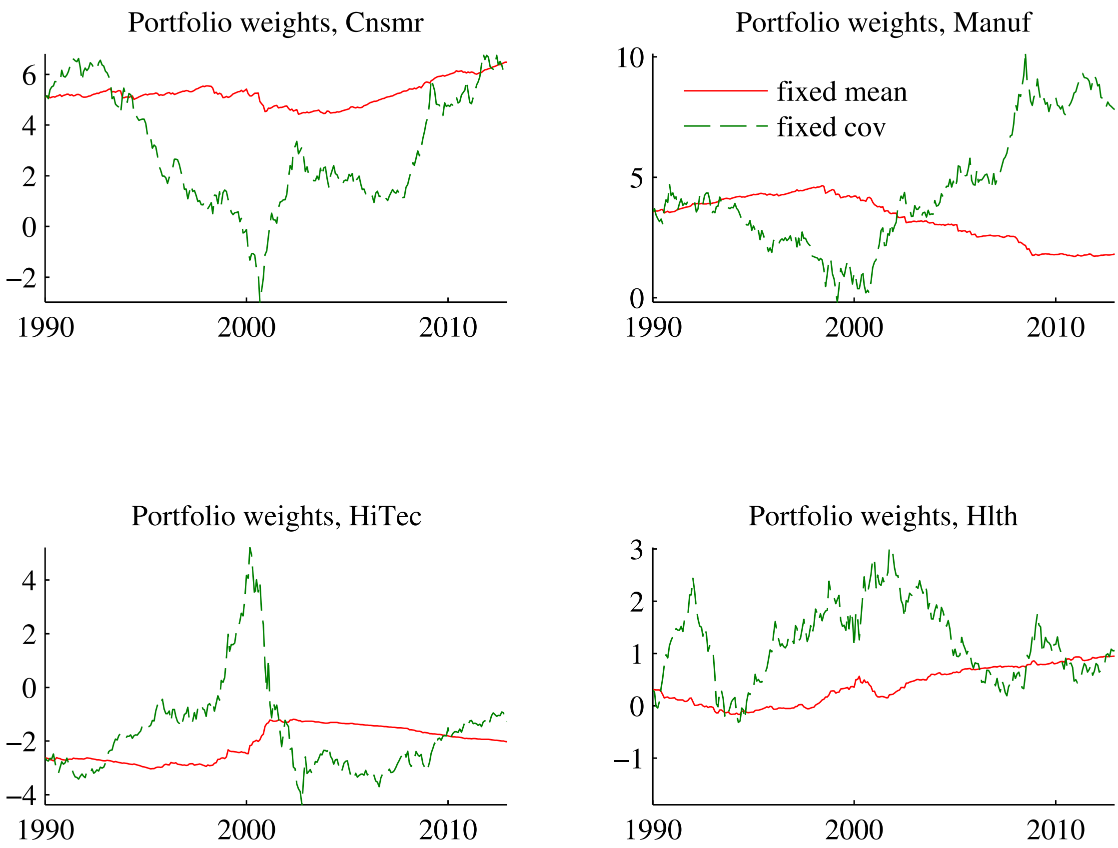


Figure 11.3: Dynamically updated portfolio weights, T-bill and 5 U.S. industries

11.2.4 A Simple Example with Time-Varying Expected Returns (Log Utility and Non-iid Returns)

A particularly simple case is when the expected excess returns are linear functions of some information variables in the $(k \times 1)$ vector z_t

$$\mu_{t+1}^e = a + bz_t, \text{ with } \mathrm{E}\, z_t = 0, \tag{11.9}$$

at the same time as the variances and covariances are constant. In this expression, a is an $n \times 1$ vector and b is an $n \times k$ matrix. Assuming that the information variables have zero means turns out to be convenient later on, but it is not a restriction (since the means are captured by a). The information variables could perhaps be the slope of the yield curve

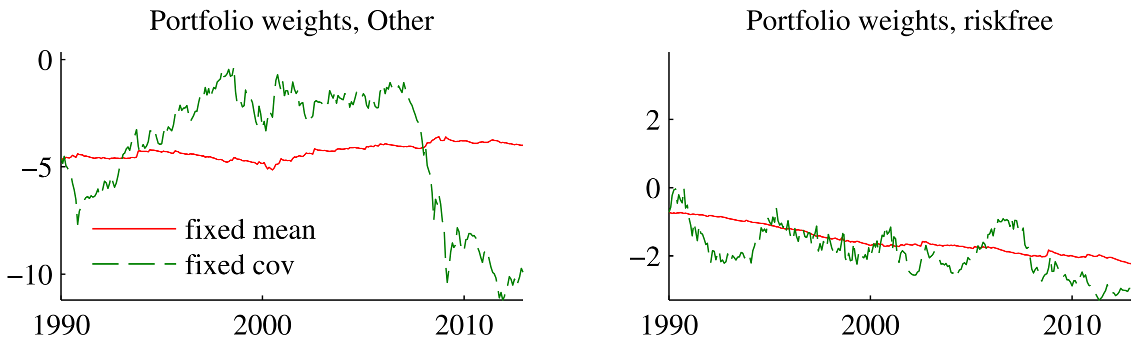


Figure 11.4: Dynamically updated portfolio weights, T-bill and 5 U.S. industries

and/or the earnings/price ratio for the aggregate stock market.

For the case with one risky asset, we get

$$v_t = \frac{\overbrace{a + bz_t}^{\mu^e_{t+1}} + \sigma^2/2}{\sigma^2}, \text{ or} \tag{11.10}$$

$$= \psi + \omega_t, \text{ with} \tag{11.11}$$

$$\psi = \frac{a + \sigma^2/2}{\sigma^2} \text{ and } \omega_t = \frac{bz_t}{\sigma^2}.$$

so the weight on the risky asset varies linearly with the information variable bz_t. (Even if there are many elements in z_t, bz_t is a scalar so it is effectively one information variable.) In the second equation, the portfolio weight is split up into the static (average) weight (ψ) and the time-varying part (ω_t). Clearly, a higher expected return implies a higher portfolio weight of the risky asset.

Similarly, for the case with many risky assets we get

$$v_t = \Sigma^{-1}\overbrace{(a + bz_t)}^{\mu^e_{t+1}} + \Sigma^{-1}\sigma^2/2, \text{ or} \tag{11.12}$$

$$= \psi + \omega_t, \text{ with} \tag{11.13}$$

$$\psi = \Sigma^{-1}(a + \sigma^2/2) \text{ and } \omega_t = \Sigma^{-1}bz_t.$$

See Figure 11.5 for an illustration (based on Example 11.3). The figure shows the basic properties for the returns, the optimal portfolios and their location in a traditional mean-std figure. In this example, z_t can only take on two different values with equal probability: -1 or 1. The figure shows one mean-variance figure for each state—and the portfolio is clearly on them. However, the portfolio is *not* on the unconditional mean-variance figure (where the means and covariance matrix are calculated by using both states).

Example 11.3 *(Dynamic portfolio weights when z_t is a scalar that only takes on the values -1 and 1, with equal probabilities) The expected excess returns are*

$$\mu^e_{t+1} = \begin{cases} a - b & \text{when } z_t = -1 \\ a + b & \text{when } z_t = 1. \end{cases}$$

The portfolio weights on the risky assets (11.13) are then

$$v_t = \begin{cases} \Sigma^{-1}(a + \sigma^2/2) - \Sigma^{-1}b & \text{when } z_t = -1 \\ \Sigma^{-1}(a + \sigma^2/2) + \Sigma^{-1}b & \text{when } z_t = 1. \end{cases}$$

Example 11.4 *(One risky asset) Suppose there is one risky asset and $a = 1, b = 2, k = 3/4, \sigma^2 = 1$,, then Example 11.3 gives*

μ^e_{t+1}	v_t	
-1	$-4/3$	*in low state*
3	4	*in high state*

Example 11.5 *(Numerical values for Example 11.3). Suppose we have three assets with*

$$\text{Cov}\left(\begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}\right) = \begin{bmatrix} 1.19 & 0.32 & 0.24 \\ 0.32 & 0.81 & 0.02 \\ 0.024 & 0.02 & 0.23 \end{bmatrix}/100,$$

and

$$\mu^e_{-1} = \begin{bmatrix} -0.41 \\ -0.29 \\ -0.07 \end{bmatrix}/100 \text{ and } \mu^e_1 = \begin{bmatrix} 0.63 \\ 0.43 \\ 0.21 \end{bmatrix}/100,$$

In this case, the portfolio weights are

$$v_{-1} \approx \begin{bmatrix} 0.112 \\ 0.094 \\ 0.065 \end{bmatrix} \text{ and } v_1 \approx \begin{bmatrix} 0.709 \\ 0.736 \\ 0.610 \end{bmatrix}.$$

Example 11.6 *(Details on Figure 11.5) To transfer from the log returns to the mean and std of net returns, the following result is used: if the vector $x \sim N(\mu, \sigma^2)$ and $y =$ $\exp(x)$, then* $\mathrm{E}\, y_i = \exp(\mu_{ii} + \sigma_{ii}/2)$ *and* $\mathrm{Cov}(y_i, y_j) = \exp\left[\mu_i + \mu_j + (\sigma_{ii} + \sigma_{jj})/2\right]\left[\exp(\sigma_{ij}) - 1\right]$.

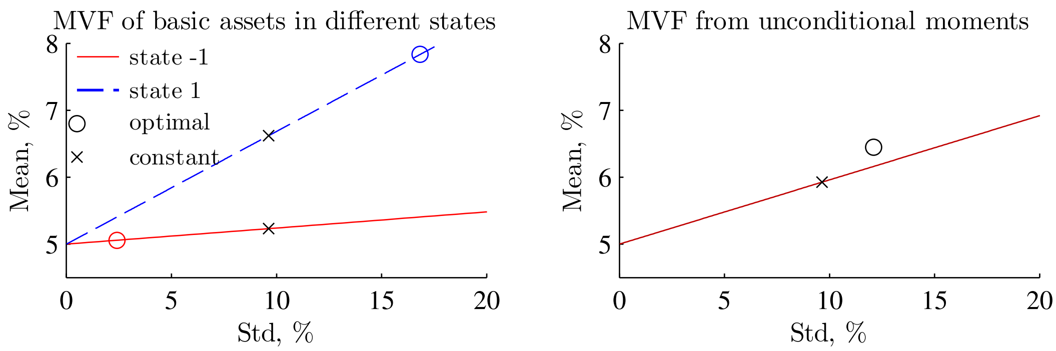


Figure 11.5: Portfolio choice, two different states

11.3 Optimal Portfolio Choice: CRRA Utility and non-iid Returns

11.3.1 Basic Setup

An important feature of the portfolio choice based on the logarithmic utility function is that it is *myopic* in the sense that it only depends on the distribution of next period's return, not on the distribution of returns further into the future. Hence, short-run and long-run investors choose the same portfolios—as discussed before. This property is special to the logarithmic utility function.

With a utility function with a constant relative risk aversion (CRRA) different from one, today's portfolio choice would also depend on distribution of returns in $t+2$ and onwards. In particular, it would depend on how the (random) returns in $t+1$ are correlated with changes (in $t+1$) of expected returns and volatilities of returns in $t+2$ and onwards. This is *intertemporal hedging*.

In this case, the optimization problem is tricky, so I will illustrate it by using a simple model. As in Campbell and Viceira (1999), suppose there is only one risky asset and let the (scalar) information variable be an AR(1)

$$z_t = \phi z_{t-1} + \eta_t, \tag{11.14}$$

where η_t is iid$N(0, \sigma_\eta^2)$. In addition, I assume that the expected return follows (11.9) but with $b = 1$ (to simplify the algebra)

$$\mu_{t+1}^e = a + z_t. \tag{11.15}$$

Combine the time series processes (11.14) and (11.15) to get the following expression for the excess return

$$r_{t+1}^e = r_{t+1} - r_f = a + z_t + u_{t+1}, \tag{11.16}$$

where u_{t+1} is iid$N(0, \sigma^2)$. Clearly, the conditional variance of the return is $\mathrm{Var}_t(r_{t+1}^e) = \mathrm{Var}(u_{t+1}) = \sigma^2$. This innovation to the return is allowed to be correlated with the shock to the future expected return, η_{t+1}, $\mathrm{Cov}(u_{t+1}, \eta_{t+1}) = \sigma_{u\eta}$. For instance, a negative correlation could be interpreted as a mean-reversion of the asset price level: a temporary positive return is followed by lower future (expected) returns.

Remark 11.7 *(*How to estimate (11.14) and (11.16)). First, regress the excess returns on some information variables z_t^*: $r_{t+1} - r_f = a^* + b^* z_t^* + u_{t+1}$. Second, define $z_t = b^*(z_t^* - \mathrm{E}\, z_t^*)$. Then, a regression of the return on z_t gives a slope coefficient of one as in (11.16). Third, estimate an AR(1) on z_t as in (11.14). Fourth and finally, estimate the covariance matrix of the residuals from the last two regressions.*

It is important to realize that the unconditional and conditional autocovariances differ markedly

$$\mathrm{Cov}(r_{t+1}^e, r_{t+2}^e) = \phi\, \mathrm{Var}(z_t) + \sigma_{u\eta} \tag{11.17}$$

$$\mathrm{Cov}_t(r_{t+1}^e, r_{t+2}^e) = \sigma_{u\eta}. \tag{11.18}$$

This shows that the unconditional autocovariance of the return can be considerable at the same time as the conditional autocovariance may be much smaller. It is the latter

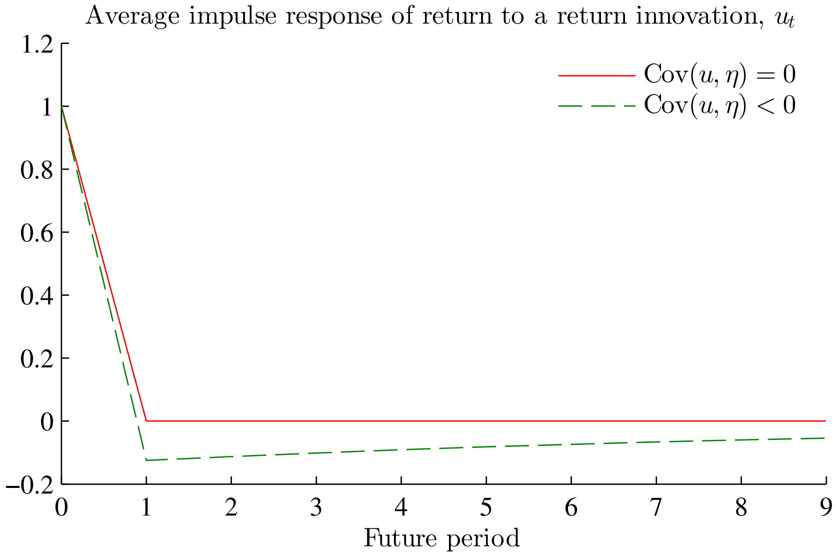


Figure 11.6: Average impulse response of the return to changes in u_0, two different cases

than matters for the portfolio choice. For instance, it is possible that the unconditional autocovariance is zero (in line with empirical evidence), while the conditional covariance is negative.

Figure 11.6 shows the impulse response function (the forecast based on current information) of a shock to the temporary part of the return (u) under two different assumptions about how this temporary part is correlated with the mean return for the next period return. When they are uncorrelated, then a shock to the temporary part of the return is just a "blip." In contrast, when today's return surprise indicates poor future returns (a negative covariance), then the impulse response function is positive (unity) in the initial period, but then negative for a prolonged period (since the expected return, $a + z_t$, is autocorrelated).

Proof. (of (11.17)–(11.18)) The unconditional covariance is

$$\begin{aligned}
\text{Cov}(r^e_{t+1}, r^e_{t+2}) &= \text{Cov}(z_t + u_{t+1}, \phi z_t + \eta_{t+1} + u_{t+2}) \\
&= \phi \,\text{Var}(z_t) + \sigma_{u\eta},
\end{aligned}$$

since $z_t + u_{t+1}$ is uncorrelated with $\eta_{t+1} + u_{t+2}$. The conditional covariance is

$$\begin{aligned} \operatorname{Cov}_t(r^e_{t+1}, r^e_{t+2}) &= \operatorname{Cov}_t(z_t + u_{t+1}, \phi z_t + \eta_{t+1} + u_{t+2}) \\ &= \sigma_{u\eta}, \end{aligned}$$

since z_t is known in t and u_{t+1} is uncorrelated with u_{t+2}. It is also straightforward to show that the unconditional variance is

$$\begin{aligned} \operatorname{Var}(r^e_{t+1}) &= \operatorname{Cov}(z_t + u_{t+1}, z_t + u_{t+1}) \\ &= \operatorname{Var}(z_t) + \operatorname{Var}(u_t), \end{aligned}$$

since z_t and u_{t+1} are uncorrelated. The conditional variance is

$$\begin{aligned} \operatorname{Var}_t(r^e_{t+1}) &= \operatorname{Cov}(z_t + u_{t+1}, z_t + u_{t+1}) \\ &= \operatorname{Var}(u_t), \end{aligned}$$

since z_t is known in t. ■

To solve the maximization problem, notice that if the log portfolio return, $r_p = \ln(1 + R_p)$, is normally distributed, then maximizing $\mathrm{E}(1 + R_p)^{1-\gamma}/(1-\gamma)$ is equivalent to maximizing

$$\mathrm{E}\, r_p + (1-\gamma)\operatorname{Var}(r_p)/2, \tag{11.19}$$

where r_p is the log return of the portfolio (strategy) over the investment horizon (one or several periods—to be discussed below).

11.3.2 One-Period Investor (Myopic Investor)

With one risky and a riskfree asset, a *one-period investor* (also called a myopic investor) maximizes

$$\mathrm{E}_t\, r_{pt+1} + (1-\gamma)\operatorname{Var}_t(r_{pt+1})/2. \tag{11.20}$$

Combine with approximate expression for r_{pt+1} (11.4) and maximize. This gives the following weight on the risky asset

$$v_t = \frac{\mu^e_{t+1} + \sigma^2/2}{\gamma\sigma^2} = \frac{a + z_t + \sigma^2/2}{\gamma\sigma^2}, \tag{11.21}$$

and the weight on the riskfree asset is $1 - v_t$. With $\gamma = 1$ (log utility), we get the same results as in (11.7). With a higher risk aversion, the weight on the risky asset is lower. Clearly, the portfolio choice depends positively on the (signal about) the expected returns. Figure 11.7 for how the portfolio weight on the risky asset depends on the risk aversion.

Example 11.8 *(Portfolio weight for one-period investor) With* $(\sigma, a, \sigma_{u\eta}, \sigma_\eta) = (0.4, 0.05, -0.4, 2)$ *and* $\gamma = 2$*, the portfolio weight in (11.21) is (on average, that is, when* $z_t = 0$*)*

$$v_t = \frac{0.05 + 0 + 0.4^2/2}{2 \times 0.4^2} \approx 0.41.$$

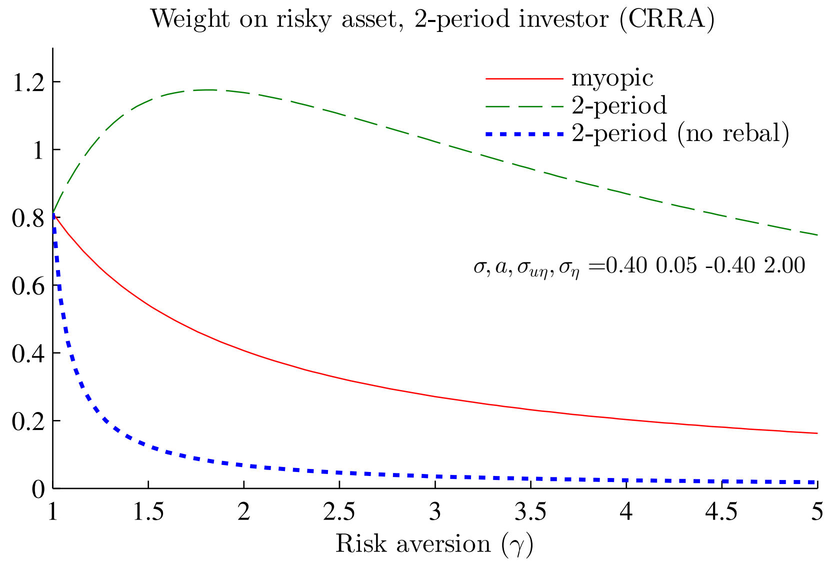


Figure 11.7: Weight on risky asset, two-period investor with CRRA utility and the possibility to rebalance

Proof. (of (11.21)). Using the approximation (11.4), we have

$$\begin{aligned} \mathrm{E}\, r_p &= r_f + v\mu^e + v\sigma^2/2 - v^2\sigma^2/2 \\ \mathrm{Var}(r_p) &= v^2\sigma^2. \end{aligned}$$

The optimization problem is therefore

$$\max_v r_f + v\mu^e + v\sigma^2/2 - v^2\sigma^2/2 + (1-\gamma)v^2\sigma^2/2,$$

so the first order condition is

$$\mu^e + \sigma^2/2 - \gamma v \sigma^2 - \gamma v \sigma^2 = 0.$$

Solve for v. ■

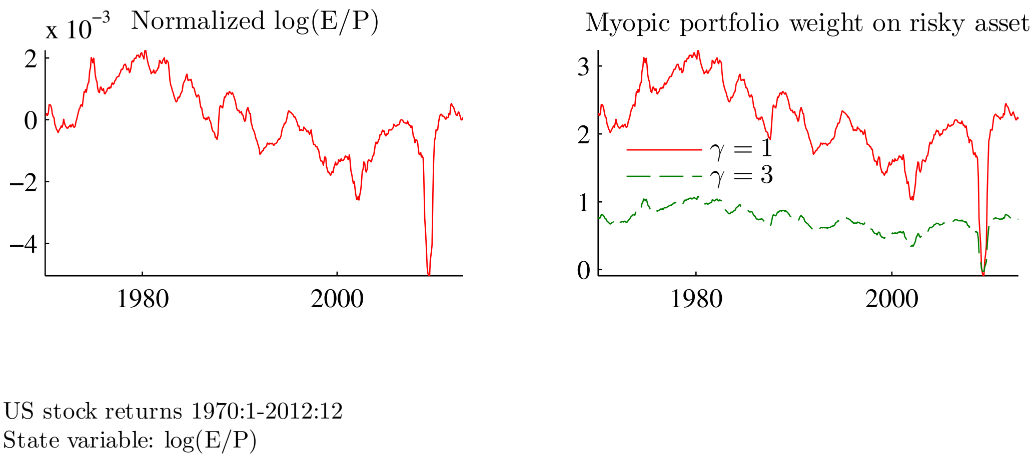


Figure 11.8: Dynamic portfolio weights

11.3.3 Two-Period Investor (No Rebalancing)

In period t, a two-period investor chooses v_t to maximize

$$\mathrm{E}_t(r_{pt+1} + r_{pt+2}) + (1-\gamma)\,\mathrm{Var}_t(r_{pt+1} + r_{pt+2})/2. \tag{11.22}$$

The solution (see Appendix) is

$$v = \frac{a + \sigma^2/2 + (1+\phi)z_t/2}{\gamma\sigma^2 - (1-\gamma)(\sigma_\eta^2/2 + \sigma_{u\eta})}. \tag{11.23}$$

Similar to the one-period investor, the weight is increasing in the signal of the average return (z_t), but there are also some interesting differences. Even if the utility function is logarithmic ($\gamma = 1$), we do not get the same portfolio choice as for the one-period investor. In particular, the reaction to the signal (z_t) is smaller (unless $\phi = 1$). The reason is that in this case, the investor commits to the same portfolio for two periods—and the movements in average returns are assumed to be mean-reverting.

There are also some important patterns on average (when $z_t = 0$). Then, $\gamma = 1$ actually gives the same portfolio choice as for the one-period investor. However, if $\gamma > 1$, and there are important shocks to the expected return, then the two-period investor puts a lower weight on the risky asset (the second term in the denominator tends to be positive). The reason is that the risky asset is more dangerous to the two-period investor since r_{pt+2} is more risky than r_{pt+1}, since r_{pt+2} can be hit by more shocks—shocks to the expected return of r_{pt+2}. In contrast, if data is iid then those shocks do not exist ($\mathrm{Var}(\eta_{t+1}) = 0$), so the two-period investor makes the same choice as the one-period investor.

One more thing is worth noticing: if $\sigma_{u\eta} < 0$, then the demand for the risky asset is higher than otherwise. This can be interpreted as a case where a temporary positive return leads to lower future (expected) returns. With this sort of mean-reversion in the price level (conditional negative autocorrelation), the risky asset is somewhat less risky to a long-run investor than otherwise. When extended to several risky assets, the result is that there us a higher demand for assets that tend to be negatively correlated with the future general investment outlook. See Figure 11.6 for an illustration of this effect and Figure 11.7 for how the portfolio weight on the risky asset depends on the risk aversion.

Example 11.9 *(Portfolio weight without rebalancing) Using the same parameters values as in Example 11.8, (11.22) is (at $z_t = 0$)*

$$v = \frac{0.05 + 0.4^2/2 + 0}{2 \times 0.4^2 - (1-2)(2^2/2 - 0.4)} \approx 0.07$$

11.3.4 Two-Period Investor (with Rebalancing)

It is more reasonable to assume that the two-period investor can rebalance in each period. Rewrite (11.22) as

$$\mathrm{E}_t\, r_{pt+1} + \mathrm{E}_t\, r_{pt+2} + (1-\gamma)[\mathrm{Var}_t(r_{pt+1}) + \mathrm{Var}_t(r_{pt+2}) + 2\,\mathrm{Cov}_t(r_{pt+1}, r_{p2+1})]/2, \tag{11.24}$$

and notice that the investor (in period t) can affect only those terms that involve r_{pt+1} (as the portfolio will be rebalanced in $t+1$). He/she therefore maximizes

$$\mathrm{E}_t\, r_{pt+1} + (1-\gamma)[\mathrm{Var}_t(r_{pt+1}) + 2\,\mathrm{Cov}_t(r_{pt+1}, r_{p2+1})]/2. \tag{11.25}$$

The maximization problem is the same as for a one-period investor (11.20) if returns are iid (so the covariance is zero), or if $\gamma = 1$.

Otherwise, the covariance term will influence the portfolio choice in t. The difference to the no-rebalancing case is that the investor in t takes into account that r_{pt+2} will be generated by a portfolio with the weights of a one-period investor

$$v_{t+1} = \frac{a + z_{t+1} + \sigma^2/2}{\gamma\sigma^2}. \tag{11.26}$$

(This is the same as (11.21) but with the time subscripts advanced one period). This affects both how the signal about future average returns (z_t) and the risk are viewed. The solution is (a somewhat messy expression, see Appendix for a proof)

$$v_t = \frac{a + z_t + \sigma^2/2}{\gamma\sigma^2} + \frac{1-\gamma}{\gamma\sigma^2}\frac{2\gamma - 1}{\gamma^2\sigma^2}\left(a + \sigma^2/2 + \phi z_t\right)\sigma_{u\eta}. \tag{11.27}$$

See Figure 11.7 for how the portfolio weight on the risky asset depends on the risk aversion and for a comparison with the cases of myopic portfolio choice and and no rebalancing.

As before, the portfolio choice depends positively on the expected return (as signalled by z_t). But, there are several other results. First, when $\gamma = 1$ (log utility), then the portfolio choice is the same as for the one-period investor (for any value of z_t). Second, when $\sigma_{u\eta} = \mathrm{Var}_t(u_{t+1}, \eta_{t+1}) = 0$, then the second term drops out, so the two-period investor once again picks the same portfolio as the one-period investor does. Third, $\gamma > 1$ combined with $\sigma_{u\eta} < 0$ increases (on average, $z_t = 0$) the weight on the risky asset—similar to the case without rebalancing. In this case, the second term of (11.27) is positive. That is, there is a positive extra demand (in t) for the risky asset: such an asset tends to pays off in $t+1$ (since $u_{t+1} > 0$, which only affects the return in $t+1$, not in subsequent periods) when the overall investment prospects for $t+2$ become worse (μ^e_{t+2} is low since η_{t+1} and thus z_{t+1} tends to be low when u_{t+1} is high and $\sigma_{u\eta} < 0$). In this case, the return in $t+1$, driven by the temporary shock u_{t+1}, partially hedges investment outlook in $t+1$ (that is, the distribution of the portfolio returns in $t+2$). The key to getting intertemporal hedging is thus that the temporary movements in the return partially offset future movements in the investment outlook.

To get a better understanding of the dynamic hedging, suppose again that we have a positive shock to the return in $t+1$, that is, $u_{t+1} > 0$. This clearly benefit all investors, irrespective of whether they are can rebalance or not. However, the investor who can rebalance in $t+1$ has advantage. His portfolio weight in $t+1$ (when he's a one-period investor) is given by (11.26), which depends on z_{t+1}. Knowing u_{t+1} does not tell us exactly what z_{t+1} is since the latter depends on the shock η_{t+1} (see (11.14)). However, we know that

$$\mathrm{E}(z_{t+1}|z_t, u_{t+1}) = \phi z_t + \mathrm{E}(\eta_{t+1}|u_{t+1}) = \phi z_t + \frac{\sigma_{u\eta}}{\sigma^2} u_{t+1}, \tag{11.28}$$

where $\sigma_{u\eta}/\sigma^2$ is the (population) regression coefficient from regressing η_{t+1} on u_{t+1}. (This follows from the standard properties of bivariate normally distributed variables.)

Therefore, the conditional expected one-period portfolio weight (11.26)

$$\mathrm{E}(v_{t+1}|z_t, u_{t+1}) = \frac{a + \phi z_t + (\sigma_{u\eta}/\sigma^2)u_{t+1} + \sigma^2/2}{\gamma\sigma^2}. \tag{11.29}$$

When $\sigma_{u\eta} < 0$, then a positive u_{t+1} (good for the return in $t+1$, but signalling poor expected returns in $t+2$) is on average followed by a lower weight (v_{t+1}) on the risky asset than otherwise. See Figure 11.9.

This shows that an investor who can rebalance can enjoy the upside (in $t+1$) without having to suffer the likely downside (in $t+2$). Conversely, when he suffers a downside in $t+1$, then he can enjoy the likely upside in $t+2$. Overall, this makes the risky asset more attractive than otherwise.

Example 11.10 *(Portfolio weight with rebalancing) Using the same parameters values as in Example 11.8, (11.27) is (at $z_t = 0$)*

$$\begin{aligned} v_t &= \frac{0.05 + 0 + 0.4^2/2}{2 \times 0.4^2} + \frac{1-2}{2 \times 0.4^2}\frac{2 \times 2 - 1}{2^2 \times 0.4^2}\left(0.05 + 0.4^2/2 + 0\right)(-0.4) \\ &\approx 0.41 + 0.76 = 1.17. \end{aligned}$$

Consider a positive shock to the return in $t+1$, for instance, $u_{t+1} = 0.1$ so $r^e_{t+1} = 0.05 + 0 + 0.1 = 0.15$. From (11.28), we have

$$\mathrm{E}(z_{t+1}|z_t, u_{t+1}) = 0 + \frac{-0.4}{0.4^2} \times 0.1 = -0.25,$$

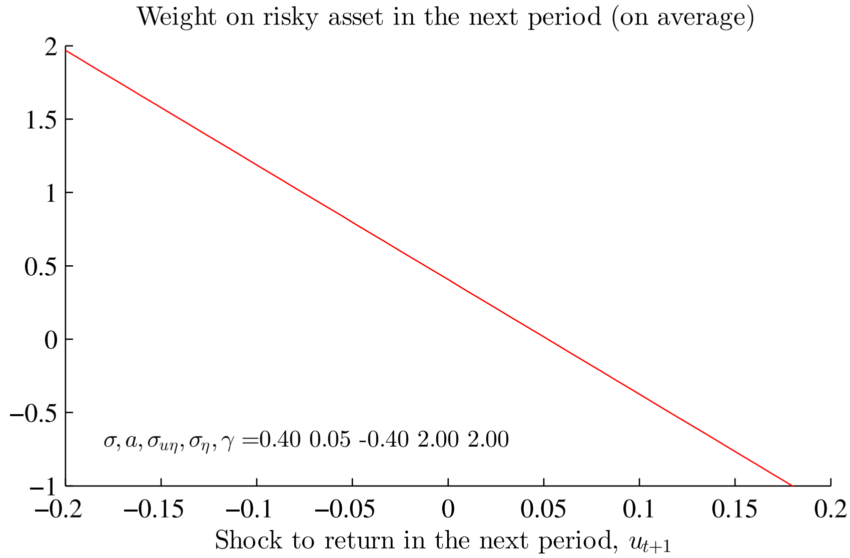


Figure 11.9: Average portfolio weight v_{t+1} as a function of u_{t+1}

so the one-period portfolio weight (11.29) is (on average, conditional on $u_{t+1} = 0.1$)

$$\mathrm{E}(v_{t+1}|z_t = 0, u_{t+1} = 0.1) = \frac{0.05 + (-0.25) + 0.4^2}{2 \times 0.4^2} = -0.375.$$

This is negative since the expected return for $t + 2$ is negative.

While this simplified case only uses one risky asset, it is important to understand that this intertemporal hedging is *not* about that a particular asset hedging the changes in its own return distribution. Indeed, if the outlook for a particular asset becomes worse, the investor could always switch out of it. Instead, the key effect depends on how a particular asset hedges the movements in tomorrow's optimal portfolio—that is, tomorrow's overall investment outlook.

11.4 Performance Measurement with Dynamic Benchmarks*

Reference: Ferson and Schadt (1996), Dahlquist and Söderlind (1999)

Traditional performance tests typically rely on the alpha from a CAPM regression. The benchmark in the evaluation is then a fixed portfolio consisting of assets that are correctly priced by the CAPM (obeys the beta representation). It often makes sense to use a more demanding benchmark—by including managed portfolios.

Let $v(z)$ be a vector of portfolio weights that potentially depend on the information variables in z. The return on such a portfolio is

$$R_{pt} = v(z)'R_t + [1 - \mathbf{1}'v(z)]R_f = v(z)'R_t^e + R_f. \tag{11.30}$$

However, without restrictions on $v(z)$ it is impossible to sort out what sort of strategies that would be assigned neutral performance by a particular (multi-factor) model. Therefore, assume that $v(z)$ are linear in the K information variables

$$v(z_{t-1}) = \underbrace{d}_{N\times K} \underbrace{z_{t-1}}_{K\times 1} \tag{11.31}$$

for any $N \times K$ matrix d. For instance, when the expected returns are driven by the information variables z_t as in (11.9), then the optimal portfolio weights (for an investor with logarithmic preferences) are linear functions of the information variables as in (11.11) or (11.13).

It is clear that the portfolio return (11.30)–(11.31) can be written

$$\begin{aligned} R_{pt} &= R_t^{e\prime} v(z_{t-1}) + R_f \\ &= R_t^{e\prime} d z_{t-1} + R_f \\ &= (\operatorname{vec} d)'(z_{t-1} \otimes R_t^e) + R_f. \end{aligned} \tag{11.32}$$

Remark 11.11 *(Kronecker product) For instance, we have that if*

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, f = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}, \text{ then } z \otimes f = \begin{bmatrix} z_1 f_1 \\ z_1 f_2 \\ z_1 f_3 \\ z_2 f_1 \\ z_2 f_2 \\ z_3 f_3 \end{bmatrix}.$$

Proof. (of (11.32)) Recall the rule that $\operatorname{vec}(ABC) = (C' \otimes A) \operatorname{vec} B$. Here, notice that $R^{e\prime}dz$ is a scalar, so we can use the rule to write $R^{e\prime}dz = (z' \otimes R^{e\prime}) \operatorname{vec} d$. Transpose and recall the rule $(D \otimes E)' = D' \otimes E'$ to get $(\operatorname{vec} d)'(z \otimes R^e)$ ■

This shows that the portfolio return can involve any linear combination of $z \otimes R^e$ so the new return space is defined by these new *managed portfolios*. We can therefore think of the returns

$$\tilde{R}_t = (z_{t-1} \otimes R_t^e) + R_f \tag{11.33}$$

as the returns on new assets—which can be used to define, for instance, mean-variance frontiers.

It is not self-evident how to measure the performance of a portfolio in this case. It could, for instance, be argued that the return of the dynamic part of the portfolio is to be considered non-neutral performance. After all, this part exploits the information in the information variables z, which is potentially better than keeping a fixed portfolio. In this case, the alpha from a traditional CAPM regression

$$R_{pt}^e = \alpha + \beta R_{mt}^e + \varepsilon_{it} \tag{11.34}$$

is a good measure of performance.

Example 11.12 *(One risky asset, two states) If the two states in Example 11.4 are equally likely and the riskfree rate is 5%, then it can be shown that* $\alpha = 4.27\%$ *and* $\beta = 2.4$.

On the other hand, it may also be argued that a dynamic trading rule that investors can easily implement themselves should be assigned neutral performance. This can be done by changing the "benchmark" portfolio from being just the market portfolio to include managed portfolios. As an example, we could use the intercept from the following "dynamic CAPM" (or "conditional CAPM") as a measurement of performance

$$\begin{aligned} R_{pt}^e &= \alpha + (\beta + \gamma z_{t-1})\, R_{mt}^e + \varepsilon_t \\ &= \alpha + \beta R_{mt}^e + \gamma z_{t-1} R_{mt}^e + \varepsilon_t. \end{aligned} \tag{11.35}$$

where the second term are the dynamic benchmarks that capture the effect of time-varying portfolio weights. In fact, (11.35) would assign neutral performance ($\alpha = 0$) to any pure "market timing" portfolio (constant relative weights in the sub portfolio of risky assets, but where the split between riskfree and risky assets change).

Remark 11.13 *In a multi-factor model we could use the intercept from*

$$R_{pt}^e = \alpha + \beta f_t + \gamma (z_{t-1} \otimes f_t) + \varepsilon_t,$$

where f_t is a vector of factors (excess returns on some portfolios), where $\otimes$ is the Kronecker product.

11.4.1 A Simple Example with Time-Varying Expected Returns

To connect the performance evaluation in (11.34) and (11.35) to the optimal dynamic portfolio strategy (11.13), suppose the optimal strategy is a pure "market timing" portfolio. This happens when the expected returns (11.9) are modelled as

$$\mu^e_{t+1} = a + bz_t, \text{ with } b = c(a + \sigma^2/2), \tag{11.36}$$

where c is some scalar constant, while a and σ^2 are vectors. This gives the portfolio weights (11.13)

$$v_{t-1} = \psi + \underbrace{\psi c z_{t-1}}_{\omega_t} = \psi(1 + cz_{t-1}), \tag{11.37}$$

where ψ is defined in (11.13). There are constant relative weights in the sub portfolio of risky assets, but the split between the risky assets (the vector v_{t-1}) and riskfree (the scalar $1 - \mathbf{1}'v_{t-1}$) and change as z_{t-1} does: market timing.

Proof. (of (11.37)) Use $b = c(a + \sigma^2/2)$ from (11.36) in (11.13)

$$\begin{aligned} \psi &= \Sigma^{-1}(a + \sigma^2/2) \\ \omega_t &= \Sigma^{-1}(a + \sigma^2/2)cz_t = \psi c z_t. \end{aligned}$$

■

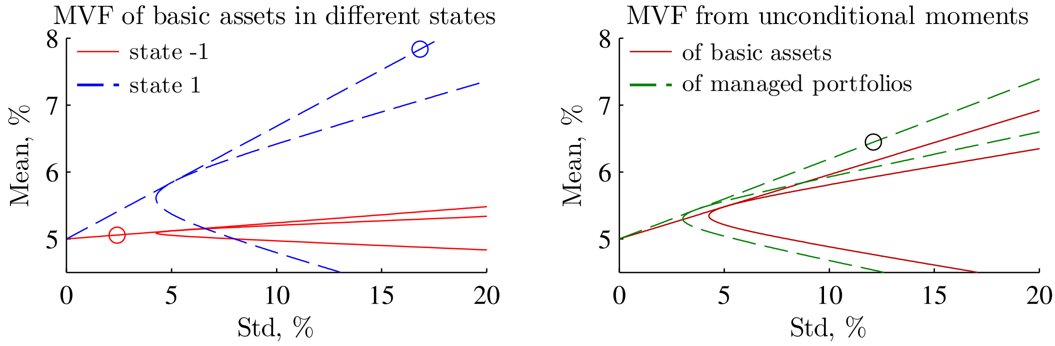
With these portfolio weights, the excess return on the portfolio is

$$R^e_{pt} = \psi' R^e_t(1 + cz_{t-1}). \tag{11.38}$$

First, consider using the intercept (α) from the the CAPM regression (11.34) as a measure of performance. If the market portfolio is the tangency portfolio (for instance, we could assume that the rest of the market do static MV optimization so the market equilibrium satisfies CAPM), then the static part of the return (11.38), $\psi' R^e_t$, will be assigned neutral performance. The dynamic part, $\psi' c z_{t-1} R^e_t$, is different: it is like the return on a new asset—which does not satisfy CAPM. It is therefore likely to be assigned a non-neutral performance.

Second, consider using the intercept from the *dynamic* CAPM regression (11.35) as a measure of performance. As before, the static part of the return should be assigned neutral performance (as the market/tangency portfolio is one of the regressors). In this case, also the dynamic part of the portfolio is likely to be assigned neutral performance (or close to it). This is certainly the case when the static portfolio weights, ψ, are proportional weights in the market portfolio. Then, the $z_{t-1}R^e_{mt}$ term in dynamic CAPM regression (11.35) exactly matches the $\psi' R^e_t z_{t-1}$ part of the return of the dynamic strategy (11.38).

See Figure 11.5 for an illustration (based on Example 11.3). Since, the portfolio is *not* on the unconditional mean-variance figure, it does not have a zero alpha when regressed against the tangency (as a proxy for the "market") portfolio. (All the basic assets do, by construction, have zero alphas.) However, it does have a zero alpha when regressed on (R_m, zR_m).



Returns:	asset 1	asset 2	asset 3
ER, state -1	5.1	5.2	5.1
ER, state 1	5.9	6.3	5.4
Std(R)	10.9	9.0	4.8

The states have equal probabilities

Correlation matrix:

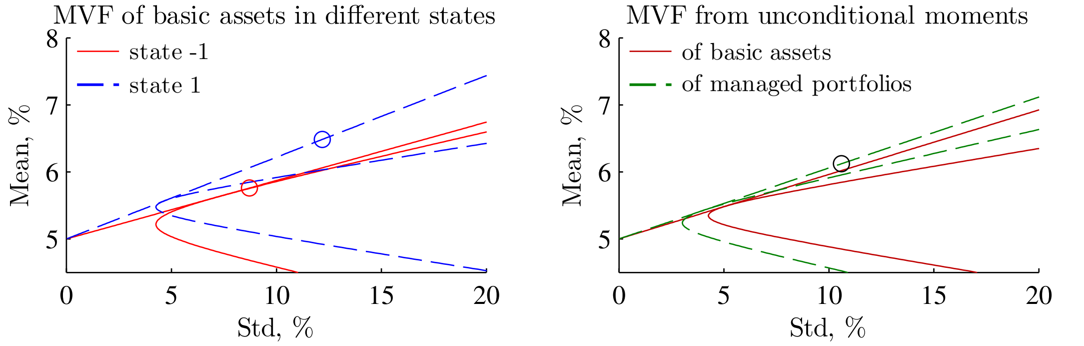
1.00	0.33	0.45
0.33	1.00	0.05
0.45	0.05	1.00

Portfolio weights:	ψ	ω_{-1}/ψ	ω_1/ψ
Asset 1	-0.03	-0.75	0.75
Asset 2	0.91	-0.75	0.75
Asset 3	1.03	-0.75	0.75

Alpha against:	R_m	$(R_m, \tilde{x}R_m)$	tangency
Asset 1	0.00	0.00	0.00
Asset 2	0.00	0.00	0.00
Asset 3	0.00	0.00	0.00
DynamicP	0.52	0.00	0.00

Figure 11.10: Portfolio choice, two different states where market timing is optimal

However, dynamic portfolio choices that are more complicated than the market timing strategy in (11.37) would not necessarily be assigned neutral performance in (11.35). However, also such strategies could be assigned a neutral performance—if we augmented



Returns:	asset 1	asset 2	asset 3
ER, state -1	5.1	5.8	5.1
ER, state 1	5.9	5.8	5.4
Std(R)	10.9	9.0	4.8

The states have equal probabilities

Correlation matrix:

1.00	0.33	0.45
0.33	1.00	0.05
0.45	0.05	1.00

Portfolio weights:	ψ	ω_{-1}/ψ	ω_1/ψ
Asset 1	-0.03	7.11	-7.11
Asset 2	0.91	0.12	-0.12
Asset 3	1.03	-0.57	0.57

Alpha against:	R_m	$(R_m, \tilde{x}R_m)$	tangency
Asset 1	0.00	0.00	0.00
Asset 2	0.00	0.00	0.00
Asset 3	0.00	0.00	0.00
DynamicP	0.20	0.16	0.00

Figure 11.11: Portfolio choice, two different states where market timing is not fully optimal

the number of benchmarks to properly capture the time-varying portfolio weights. In this case, this would require using $z_{t-1} \otimes R^e_t$ (where R^e_t are the returns on the original assets) as the regressors

$$R^e_{pt} = \alpha + \beta R^e_{mt} + \gamma(z_{t-1} \otimes R^e_t) + \varepsilon_t. \tag{11.39}$$

With those benchmarks all strategies where the portfolio weights on the original assets are linear in z_{t-1} would be assigned neutral performance. In practice, evaluation of mutual funds typically define a small number (perhaps 5) of returns and even fewer instruments (perhaps 2–3). The instruments are typically inspired by the literature on return predictability and often include the slope of the yield curve, the dividend yield or lagged returns.

Figures 11.10 illustrates the case when the portfolio has a zero alpha against (R_m, zR_m), while Figure 11.11 shows a case when the portfolio does not.

A Some Proofs

Proof. (of (11.23)) (This proof is a bit crude, but probably correct....) The objective is to maximize (11.24). Using (11.4) we have

$$r_{pt+1} \approx r_f + v r^e_{t+1} + v\sigma^2/2 - v^2\sigma^2/2$$
$$r_{pt+2} \approx r_f + v r^e_{t+2} + v\sigma^2/2 - v^2\sigma^2/2,$$

so

$$r_{pt+1} + r_{pt+2} \approx 2r_f + v(r^e_{t+1} + r^e_{t+2}) + v\sigma^2 - v^2\sigma^2.$$

The expected value of the two-period return is

$$\mathrm{E}_t(r_{pt+1} + r_{pt+2}) = 2r_f + v(\mu^e_{t+1} + \mathrm{E}_t\,\mu^e_{t+2}) + v\sigma^2 - v^2\sigma^2,$$

so the derivative with respect to v

$$\frac{\partial\,\mathrm{E}_t(r_{pt+1} + r_{pt+2})}{\partial v_t} = \mu^e_{t+1} + \mathrm{E}_t\,\mu^e_{t+2} + \sigma^2 - 2v\sigma^2. \tag{foc1}$$

The variance of the two-period return is

$$\mathrm{Var}_t(r_{pt+1} + r_{pt+2}) = v^2\,\mathrm{Var}_t(r^e_{t+1} + r^e_{t+2}),$$

so the derivative is

$$\frac{\partial\,\mathrm{Var}_t(r_{pt+1} + r_{pt+2})}{\partial v_t} = 2v\,\mathrm{Var}_t(r^e_{t+1} + r^e_{t+2}). \tag{foc2}$$

Combine (foc1) and (foc2) to get the first order condition

$$\begin{aligned} 0 &= \frac{\partial\,\mathrm{E}_t(r_{pt+1} + r_{pt+2})}{\partial v_t} + \frac{1-\gamma}{2}\frac{\partial\,\mathrm{Var}_t(r_{pt+1} + r_{pt+2})}{\partial v_t} \\ &= \mu^e_{t+1} + \mathrm{E}_t\,\mu^e_{t+2} + \sigma^2 - 2v\sigma^2 + (1-\gamma)v\,\mathrm{Var}_t(r^e_{t+1} + r^e_{t+2}), \end{aligned}$$

so we can solve for the portfolio weight as

$$v = \frac{\mu^e_{t+1} + \mathrm{E}_t\,\mu^e_{t+2} + \sigma^2}{2\sigma^2 - (1-\gamma)\,\mathrm{Var}_t(r^e_{t+1} + r^e_{t+2})}.$$

Recall that

$$\begin{aligned}
\mu^e_{t+1} &= a + z_t \\
\mathrm{E}_t\, \mu^e_{t+2} &= a + \mathrm{E}_t\, z_{t+1} = a + \phi z_t, \text{ so} \\
\mu^e_{t+1} + \mathrm{E}_t\, \mu^e_{t+2} &= 2a + (1+\phi) z_t.
\end{aligned}$$

Notice also that $r^e_{t+1} - \mathrm{E}_t\, r^e_{t+1} = u_{t+1}$ and that $r^e_{t+2} - \mathrm{E}_t\, r^e_{t+2} = \eta_{t+1} + u_{t+2}$,

$$\mathrm{Var}_t(r^e_{t+1} + r^e_{t+2}) = \mathrm{Var}_t(u_{t+1} + \eta_{t+1} + u_{t+2}) = \sigma^2 + \sigma^2_\eta + \sigma^2 + 2\sigma_{u\eta},$$

since $\mathrm{Cov}(u_{t+1}, u_{t+2}) = \mathrm{Cov}(\eta_{t+1}, u_{t+2}) = 0$. Combining into the expression for v gives

$$\begin{aligned}
v &= \frac{2a + (1+\phi) z_t + \sigma^2}{2\sigma^2 - (1-\gamma)(2\sigma^2 + \sigma^2_\eta + 2\sigma_{u\eta})} \\
&= \frac{a + (1+\phi) z_t/2 + \sigma^2/2}{\sigma^2 - (1-\gamma)(\sigma^2 + \sigma^2_\eta/2 + \sigma_{u\eta})} \\
&= \frac{a + (1+\phi) z_t/2 + \sigma^2/2}{\sigma^2\gamma - (1-\gamma)(\sigma^2_\eta/2 + \sigma_{u\eta})}.
\end{aligned}$$

■

Proof. (of (11.27)) (This proof is a bit crude, but probably correct....) The objective is to maximize

$$\mathrm{E}_t\, r_{pt+1} + (1-\gamma)[\mathrm{Var}_t(r_{pt+1})/2 + \mathrm{Cov}_t(r_{pt+1}, r_{p2+1})]. \tag{obj}$$

Using (11.4) we have

$$\begin{aligned}
r_{pt+1} &\approx r_f + v_t \left(r_{t+1} - r_f\right) + v_t\sigma^2/2 - v_t^2\sigma^2/2 \\
r_{pt+2} &\approx r_f + v_{t+1} \left(r_{t+2} - r_f\right) + v_{t+1}\sigma^2/2 - v_{t+1}^2\sigma^2/2.
\end{aligned}$$

The derivative with respect to v of the expected return in (obj) is

$$\frac{\partial\, \mathrm{E}_t\, r_{pt+1}}{\partial v_t} = \mu^e_{t+1} + \sigma^2/2 - v_t\sigma^2. \tag{foc1}$$

The variance term in (obj) is

$$\mathrm{Var}_t(r_{pt+1}) = v_t^2\,\mathrm{Var}_t\,(r_{t+1}) = v_t^2\sigma^2,$$

since $r_{t+1} - r_f = a + z_t + u_{t+1}$. The derivative of the variance part of (obj) is

$$\frac{1-\gamma}{2}\frac{\partial\,\mathrm{Var}_t(r_{pt+1})}{\partial v_t} = (1-\gamma)v_t\sigma^2. \tag{foc2}$$

The covariance in (obj) is

$$\begin{aligned}\mathrm{Cov}_t(r_{pt+1}, r_{p2+1}) &= v_t\,\mathrm{Cov}_t\left[u_{t+1}, v_{t+1}\left(r_{t+2} - r_f\right) + v_{t+1}\sigma^2/2 - v_{t+1}^2\sigma^2/2\right],\\ &= v_t\,\mathrm{Cov}_t(u_{t+1}, \underbrace{v_{t+1}\mu_{t+2}^e + v_{t+1}\sigma^2/2 - v_{t+1}^2\sigma^2/2}_{B}),\end{aligned} \tag{ff}$$

where the second line uses the fact that $r_{t+2} - r_f = \mu_{t+2}^e + u_{t+2}$ and that u_{t+2} is uncorrelated with u_{t+1} and v_{t+1}. There are two channels for the covariance: u_{t+1} might be correlated with the expected return, μ_{t+2}^e, or with the portfolio weight, v_{t+1}. The portfolio weight from the one-period optimization (11.21), but for $t + 1$, is

$$v_{t+1} = \frac{\tilde{a} + z_{t+1}}{\gamma\sigma^2},$$

where $\bar{a} = a + \sigma^2/2$ (this notation is only used to make the subsequent equations shorter) The B term in (ff) can then be written

$$\begin{aligned}B &= (\bar{a} + z_{t+1})\,(\bar{a} + z_{t+1})\,\frac{1}{\gamma\sigma^2}\left(1 - \frac{1}{2\gamma}\right)\\ &= \left(2\bar{a}z_{t+1} + z_{t+1}^2\right)\frac{1}{\gamma\sigma^2}\left(1 - \frac{1}{2\gamma}\right) + \text{constants}\end{aligned}$$

Since $z_{t+1} = \phi z_t + \eta_{t+1}$, we have $z_{t+1}^2 = \phi^2 z_t^2 + \eta_{t+1}^2 + 2\phi z_t\eta_{t+1}$. Dropping variables known in t, we therefore have

$$B = \left[2\,(\bar{a} + \phi z_t)\,\eta_{t+1} + \eta_{t+1}^2\right]\frac{1}{\gamma\sigma^2}\left(1 - \frac{1}{2\gamma}\right) + \text{known in } t$$

Since $\mathrm{Cov}_t\left(u_{t+1}, \eta_{t+1}^2\right) = 0$ (since they are jointly normally distributed) the covariance in (ff)

$$\mathrm{Cov}_t(r_{pt+1}, r_{p2+1}) = v_t\,(\bar{a} + \phi z_t)\,\sigma_{u\eta}\frac{1}{\gamma\sigma^2}\left(2 - \frac{1}{\gamma}\right)$$

The derivative of the covariance part of (obj) is

$$(1-\gamma)\frac{\partial\,\mathrm{Cov}_t(r_{pt+1}, r_{p2+1})}{\partial v_t} = (1-\gamma)\left(2-\frac{1}{\gamma}\right)\frac{\bar{a}+\phi z_t}{\gamma\sigma^2}\sigma_{u\eta}. \tag{foc3}$$

Combine the derivatives (foc1), (foc2) and (foc3) to the first order condition

$$\begin{aligned}
0 &= \frac{\partial\,\mathrm{E}\,r_{pt+1}}{\partial v_t} + (1-\gamma)\frac{\partial\,\mathrm{Var}_t(r_{pt+1})/2}{\partial v_t} + (1-\gamma)\frac{\partial\,\mathrm{Cov}_t(r_{pt+1}, r_{p2+1})}{\partial v_t} \\
&= (\mu^e_{t+1} + \sigma^2/2 - v_t\sigma^2) + (1-\gamma)v_t\sigma^2 + (1-\gamma)\left(2-\frac{1}{\gamma}\right)\frac{\bar{a}+\phi z_t}{\gamma\sigma^2}\sigma_{u\eta} \\
&= \mu^e_{t+1} + \sigma^2/2 - \gamma v_t\sigma^2 + (1-\gamma)\left(2-\frac{1}{\gamma}\right)\frac{\bar{a}+\phi z_t}{\gamma\sigma^2}\sigma_{u\eta} \\
&= \mu^e_{t+1} + \sigma^2/2 + (1-\gamma)\left(2-\frac{1}{\gamma}\right)\frac{\bar{a}+\phi z_t}{\gamma\sigma^2}\sigma_{u\eta} - \sigma^2\gamma v_t,
\end{aligned}$$

which can be solved as (11.27). ■

Bibliography

Campbell, J. Y., and L. M. Viceira, 1999, "Consumption and portfolio decisions when expected returns are time varying," *Quarterly Journal of Economics*, 114, 433–495.

Campbell, J. Y., and L. M. Viceira, 2002, *Strategic asset allocation: portfolio choice of long-term investors*, Oxford University Press.

Dahlquist, M., and P. Söderlind, 1999, "Evaluating portfolio performance with stochastic discount factors," *Journal of Business*, 72, 347–383.

Ferson, W. E., and R. Schadt, 1996, "Measuring fund strategy and performance in changing economic conditions," *Journal of Finance*, 51, 425–461.