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

Paul Söderlind[1]

23 December 2020

[1]University of St. Gallen. *Address:* s/bf-HSG, Unterer Graben 21, CH-9000 St. Gallen, Switzerland. *E-mail:* Paul.Soderlind@unisg.ch. Document name: Fin1All.TeX.

Contents

Warning: a few of the tables and figures are reused in later chapters. This can mess up the references, so that the text refers to a table/figure in another chapter. No worries: it is really the same table/figure. Still, I promise to fix this some day.

Chapter 1

The Basics of Return Calculations

Reference: Elton, Gruber, Brown, and Goetzmann (2014) 1–3

1.1 Portfolio Return: Definition, Mean and Variance

Many portfolio choice models centre 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 Asset Return: Definition

The *net (rate of) return* on asset i in period t is

$$R_{i,t} = \frac{V_{i,t} - V_{i,t-1}}{V_{i,t-1}} = \frac{V_{i,t}}{V_{i,t-1}} - 1, \tag{1.1}$$

where $V_{i,t}$ is the value of asset i in period t.

The gross return is

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

Example 1.1 *(Returns)*

$$R = \frac{110 - 100}{100} = 0.1 \text{ (or } 10\%\text{)}$$
$$1 + R = \frac{110}{100} = 1.1$$

Remark 1.2 *(% and bp) Recall that* 6% *means* $6/100 = 0.06$*, and* 400 *bp (basis points) means* $400/10000 = 0.04$*. Warning: if you just drop the % symbol and thus effectively* *work with* $100R$ *(in this case getting 6), then you have to be careful, in particular, when accumulating returns over time and when calculating variances.*

In many cases, the values are

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

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 1.3 *(Dividend yield ad capital gain yield)*

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

An *excess return* (compared to the riskfree rate R_f) is

$$R^e_{i,t} = R_{i,t} - R_{f,t}. \tag{1.5}$$

In this particular case the reference return is the riskfree return, but it could also be the return of some other asset.

Example 1.4 *(Excess return) If* $R_{i,t} = 0.08$ *and* $R_{f,t} = 0.01$*, then the excess return is 0.07 (7%).*

The expected value of an excess return, $\mathrm{E}\, R^e_i$, is often called the *risk premium*. The *Sharpe ratio* is $\mathrm{E}\, R^e_i / \operatorname{Std}(R^e_i)$.

Example 1.5 *(Risk premium and Sharpe ratio) If* $\mathrm{E}\, R_i = 0.1$ *and* $\operatorname{Std}(R^e_i) = 0.5$*, then Sharpe ratio is 0.2.*

Remark 1.6 *(Approximating the riskfree return) Suppose you have monthly equity returns and want to calculate excess returns. Do as follows. First, find a representative money market instrument (for instance, a T-bill or an interbank contract) with approximately one month to maturity. Second, use* $R_{f,t} \approx y_{t-1\,month}/12$ *where* $y_{t-1\,month}$ *is the* *quoted interest rate one month earlier (because this is the rate you earn/pay on a loan between t−1 month and t). The result is an approximation since interest rates are quoted in different ways (simple, effective,...) and because the maturity may not be an exact match with the investment horizon.*

It is sometimes better to work with *log returns*, especially when we compare different investment horizons for the same asset. In contrast, log returns are inconvenient when the focus is on choosing the portfolio weights: the log return of a portfolio is *not* a weighted average of the log returns of the assets in the portfolio (see Example 1.12 below).

Anwhow, a log return is defined as

$$r_{i,t} = \ln(1 + R_{i,t}), \tag{1.6}$$

which clearly equals $\ln(V_{i,t}/V_{i,t-1})$.

The corresponding excess log return is

$$r^e_{i,t} = \ln(1 + R_{i,t}) - \ln(1 + R_{f,t}). \tag{1.7}$$

Assuming we invest an equal amount in both instruments in $t - 1$ ($V_{i,t-1} = V_{f,t-1}$), the excess log return equals $\ln(V_{i,t}/V_{f,t})$. Notice that excess log return is *not* the log of the excess return. (Yes, the order of the words matters here.) Rather, it is the log of $(1 + R_{i,t})/(1 + R_{f,t})$.

Example 1.7 *(Excess log return) If* $R_{i,t} = 0.08$ *and* $R_{f,t} = 0.01$*, then the excess log return in 0.067.*

1.1.2 Cumulating Returns

If your investment in asset i in period $t = 0$ equals $V_{i,0}$, then its value in t is

$$V_{i,t} = V_{i,0}(1 + R_{i,1})(1 + R_{i,2}) \dots (1 + R_{i,t}), \tag{1.8}$$

provided all dividends have been reinvested in the same asset. We can clearly write this on recursive form as

$$V_{i,t} = V_{i,t-1}(1 + R_{i,t}). \tag{1.9}$$

See Figure 1.1 for an illustration.

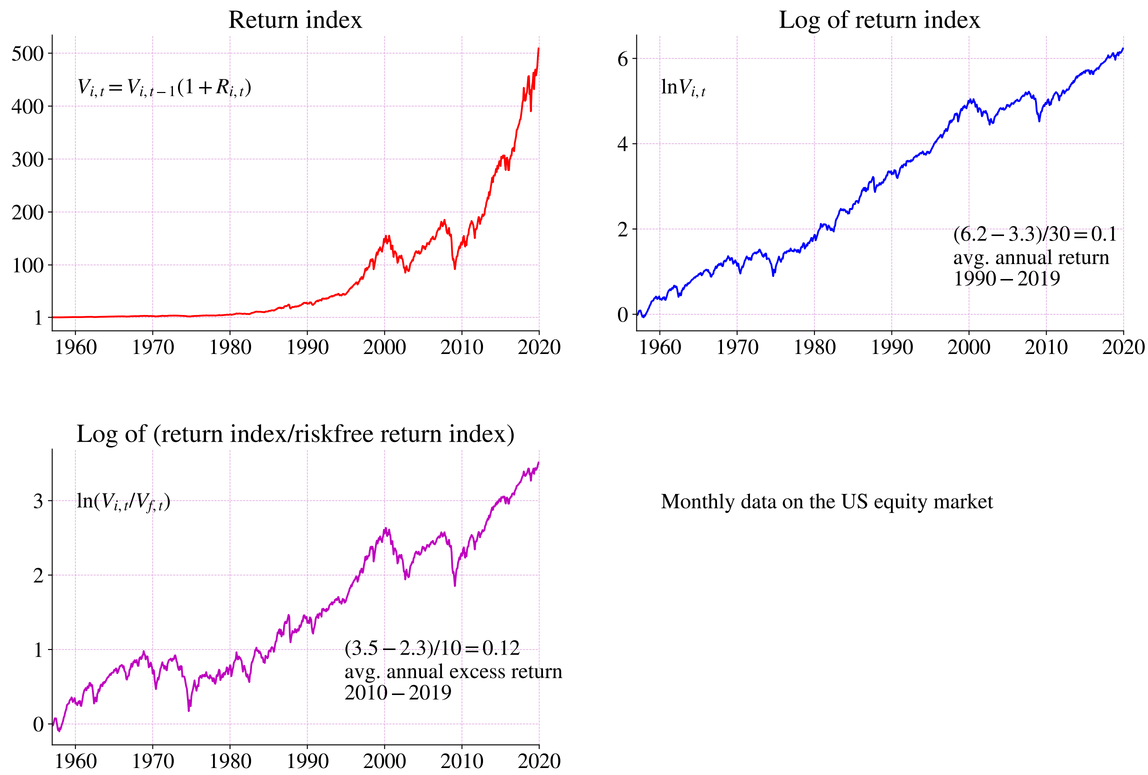


Figure 1.1: Cumulating returns

Example 1.8 *With net returns* $(R_{i,1}, R_{i.2}, R_{i,3}) = (0.2, -0.35, 0.25)$*, we get portfolio values* $V_{i,t} = (1.2, 0.78, 0.975)$ *for period* $1-3$ *(assuming* $V_{i,0} = 1$*). This is a* total return index.

Example 1.9 *The net returns in Example 1.8 are (20%,-35%,25%). Just dropping the % symbol and trying to use (1.8) as* $1(1+20)...$ *gives nonsense.*

Unfortunately, excess returns cannot be cumulated directly. Instead, you need to cumulate $R_{i,t}$ and $R_{f,t}$ separately (as in (1.9)) and then form the difference

$$V_{i,t}^{e} = V_{i,t} - V_{f,t}. \tag{1.10}$$

Sometimes the ratio $V_{i,t}/V_{f,t}$ is a preferred way of illustrating the performance of the two assets.

Remark 1.10 *(Adjusted closing price) The adjusted closing price of an asset is an index calculated as (1.9) where* $R_{i,t}$ *is the return (including dividends, splits, etc) of holding the* *asset from $t-1$ to t. This means that it is a total return index. If you have such an index, then the returns can be calculated as $R_{i,t} = V_{i,t}/V_{i,t-1} - 1$, without having to handle dividend payments separately.*

Similarly, the log value can be calculated as

$$\ln V_{i,t} = \ln V_{i,0} + r_{i,1} + r_{i,2} \ldots + r_{i,t}, \text{ so} \tag{1.11}$$

$$= \ln V_{i,t-1} + r_{i,t}. \tag{1.12}$$

You *can* cumulate excess log returns (because it is just summing). Since the initial positions are equal ($V_{i,0} = V_{f,0}$) we have

$$\ln(V_{i,t}/V_{f,t}) = (r_{i,1} + r_{i,2} + \ldots + r_{i,t}) - \left(r_{f,1} + r_{f,2} \ldots + r_{f,t}\right) \tag{1.13}$$

$$= r^e_{i,1} + r^e_{i,2} + \ldots + r^e_{i,t}, \text{ so} \tag{1.14}$$

$$= \ln(V_{i,t-1}/V_{f,t-1}) + r^e_{i,t}, \tag{1.15}$$

starting from $\ln(V_{i,0}/V_{f,0}) = 0$. Notice that the exponential function of this gives the ratio $V_{i,t}/V_{f,t}$ (not the difference).

Again, see Figure 1.1 for an illustration

1.1.3 Portfolio Return: Definition

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 $w_1, w_2, \ldots, w_n$ ($\Sigma_{i=1}^n w_i = 1$) is

$$R_{p,t} = w_1 R_{1,t} + w_2 R_{2,t} \text{ (with } n = 2) \tag{1.16}$$

$$= \sum_{i=1}^{n} w_i R_{i,t} \text{ (more generally, but assuming } \Sigma_{i=1}^n w_i = 1). \tag{1.17}$$

Example 1.11 *(Portfolio return) With the portfolio weights 0.8 and 0.2 for two assets and the returns 0.1 and 0.05 for the same assets, the portfolio has the return*

$$R_p = 0.8 \times 0.10 + 0.2 \times 0.05 = 0.09,$$

that is, 9%.

Example 1.12 *(Log portfolio return) The log portfolio return of the portfolio in Example 1.12 is* $\ln(1 + 0.09) \approx 0.086$. *Notice that this must be calculated in two steps: first* *calculate the net portfolio return and then convert to log returns. With few assets and stable returns, as in Example 1.12, the average of the log returns might give a reasonable approximation. (It can be shown that this is the first-order Taylor approximation around the point where the log returns of all assets are zero.)*

Proof. (of (1.17)) 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

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

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

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

where the last equality follows from $\sum_{i=1}^{n} w_i = 1$. Subtract 1 from both sides to get the net portfolio return (1.17). ■

Example 1.13 *(Number of assets and portfolio returns) For asset* 1 *we have* $P_{1,t-1} = 10$, $P_{1,t} = 11$ *and for asset* 2 *we have* $P_{2,t-1} = 8$, $P_{2,t} = 8.4$. *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* $R_p = \frac{218-200}{200} = 0.09$. *Compare that to (1.17) which would give*

$$R_p = 0.8 \times 0.10 + 0.2 \times 0.05 = 0.09,$$

since the two returns are 0.1 $(11/10 - 1)$ *and* 0.05 $(8.4/8 - 1)$ *respectively, and the portfolio weights are* 0.8 $(16 \times 10/200)$ *and* 0.2 $(5 \times 8/200)$ *respectively.*

The portfolio weights in (1.17) should sum to unity ($\Sigma_{i=1}^{n} w_i = 1$), but some weights could potentially be negative: *"short" positions*. Notice that a short position pays off if the asset price decreases. Clearly, some investors have very strict limits on their positions.

For instance, mutual funds can typically not shorten assets and not put more that 10% into a particular asset. In contrast, hedge funds have very few limits.

Remark 1.14 *(Short selling) How can we short sell an asset? Borrow the asset (for a fee) and sell it. A short position is profitable if the asset price decreases since then we can buy it back (to return it to the asset lender) for less than for what we sold it. If there are derivatives on the asset, then we don not need to borrow it: just issue a futures/option.*

Example 1.15 *(Return on a short position) Suppose you borrow an asset (for one month, at a fee of* 0.5*) and sell it for* 100*. One month later, you buy the asset on the market for* 90*. Your profit is thus* $100 - 90 - 0.5 = 9.5$*. Expressed in terms of the initial value of the asset, this is a return of* 9.5%*. (If you can invest the 100, this may be even higher.) In practice, you typically have to provide collateral when borrowing assets.*

A *zero-cost "portfolio"* (also called an arbitrage portfolio) is an extreme case where the initial investment is zero. This means that the investor shortens some assets (perhaps borrows) in order to invest into other (perhaps risky) assets. The return on such a portfolio is not well defined (dividing by zero...), but we can define an excess return as follows. Split up the portfolio into a "long" portfolio (those assets that we hold positive amounts of) and a "short" portfolio (those assets that we hold negative amounts of). Clearly, the (initial) value of the long portfolio must equal the value of the short portfolio (so the overall portfolio is a zero-cost portfolio).

To be precise, let w_i^L be the portfolio weight on asset i in the long portfolio and let w_i^S be the portfolio weight in the short portfolio. Clearly, $w_i^L \geq 0$ and $w_i^S \geq 0$ and when one of them is positive then the other is zero ($w_i^L w_i^S = 0$).

Example 1.16 *(Zero-cost portfolio) Suppose you invest 40 in asset 1, 60 in asset 2 and -100 in asset 3. The total investment is zero. We then have* $w^L = (0.4, 0.6, 0)$ *and* $w^S = (0, 0, 1)$.

Define the returns on the long and short portfolios as

$$R_p^L = \sum_{i=1}^n w_i^L R_i \quad (1.18)$$

$$R_p^S = \sum_{i=1}^n w_i^S R_i. \quad (1.19)$$

We can then consider an "excess return" of the total portfolio as

$$R_p^e = R_p^L - R_p^S = \sum_{i=1}^n (w_i^L - w_i^S) R_i. \quad (1.20)$$

Example 1.17 *(Excess return of a zero-cost portfolio) If the returns of the assets in Example 1.16 are $R_1 = 10\%$, $R_2 = -1\%$ and $R_3 = 2\%$, then the excess return is*

$$0.4 \times 0.1 + 0.6 \times (-0.01) - 0.02 = 0.014.$$

Remark 1.18 *(A broader definition of excess returns**) *The definition of the excess return of a zero-cost portfolio discussed above uses portfolio weights that sum to unity ($\Sigma w_i^L = 1$ and $\Sigma w_i^S = 1$), which is often a natural choice. However, another convention is used in some cases: the "excess return" of a zero cost portfolio is just its payoff (profit).*

1.1.4 Portfolio Return: Expected Value and Variance

Remark 1.19 *(Expected value and variance of a linear combination) Recall that if w_1 and w_2 are two constants, while the returns R_1 and R_2 are random variables, then*

$$\mathrm{E}(w_1 R_1 + w_2 R_2) = w_1 \mathrm{E}\, R_1 + w_2 \mathrm{E}\, R_2, \text{ and}$$
$$\mathrm{Var}(a R_1 + b R_2) = a^2 \sigma_{11} + w_2^2 \sigma_{22} + 2 w_1 w_2 \sigma_{12},$$

where $\sigma_{ij} = \mathrm{Cov}(R_i, R_j)$, and $\sigma_{ii} = \mathrm{Cov}(R_i, R_i) = \mathrm{Var}(R_i)$. Notice: σ_{ii} here denotes the variance of return i and σ_{ij} the covariance between i and j.

Remark 1.20 *(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, σ_{ii} or $\mathrm{Var}(R_i)$ 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.*

Remark 1.21 *(Scaling of returns by 100) Let R be the return and assume that you work with $\widetilde{R} = 100R$. Suppose the return R_i has the mean μ_i and variance σ_i^2. Then, $\tilde{R}_i$ has the mean $100\mu_i$, variance $100^2\sigma_i^2$, standard deviation $100\sigma_i$ and the covariance with $\widetilde{R}_j$ is $100^2\,\mathrm{Cov}(R_i, R_j)$ if both returns are scaled by 100. This can cause problems when trading off means and variances. However, it works well when comparing means and standard deviations (for instance, the Sharpe ratio is a mean divided by a standard deviation). In a regression, $R_i = \alpha + \beta R_j + \varepsilon$, the slope is unaffected when both returns are scaled by 100, but the intercept is scaled by 100.*

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

$$\mathrm{E}\, R_p = w_1 \mathrm{E}\, R_1 + w_2 \mathrm{E}\, R_2 \text{ (with } n = 2) \tag{1.21}$$

$$= \sum_{i=1}^{n} w_i \mathrm{E}\, R_i \text{ (more generally).} \tag{1.22}$$

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

$$\mathrm{Var}(R_p) = w_1^2\sigma_{11} + w_2^2\sigma_{22} + 2w_1w_2\sigma_{12} \text{ (with } n = 2) \tag{1.23}$$

$$= \sum_{i=1}^{n} w_i^2\sigma_{ii} + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} w_i w_j \sigma_{ij} \text{ (more generally).} \tag{1.24}$$

In matrix form we have

$$\mathrm{E}\, R_p = w' \mathrm{E}\, R \text{ and} \tag{1.25}$$

$$\mathrm{Var}(R_p) = w' \Sigma w. \tag{1.26}$$

Example 1.22 *(Expected value and variance of portfolio return) Let the portfolio weights be $w = [0.8, 0.2]$. Assume the following the expected values and covariance matrix for the returns:* $\mathrm{E}\, R = \begin{bmatrix} 9 \\ 6 \end{bmatrix} /100$ *and* $\Sigma = \begin{bmatrix} 256 & 96 \\ 96 & 144 \end{bmatrix} /100^2$*, we have*

$$\mathrm{E}\, R_p = \begin{bmatrix} 0.8 & 0.2 \end{bmatrix} \begin{bmatrix} 9 \\ 6 \end{bmatrix} \frac{1}{100} = 0.084,$$

$$\mathrm{Var}(R_p) = \begin{bmatrix} 0.8 & 0.2 \end{bmatrix} \begin{bmatrix} 256 & 96 \\ 96 & 144 \end{bmatrix} \frac{1}{100^2} \begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix} \approx 0.020, \textit{ and}$$

$$\mathrm{Std}(R_p) \approx 0.142.$$

Example 1.23 *(Expected value and variance of portfolio return 2) Redo the example above, but when returns are actually multiplied by 100:*

$$\mathrm{E}\, R_p = 8.4$$

$$\mathrm{Var}(R_p) \approx 200, \textit{ and}$$

$$\mathrm{Std}(R_p) \approx 14.2.$$

Remark 1.24 *(Details on the matrix form) With two assets, we have the following:*

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

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

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

1.1.5 Some Practical Remarks

Remark 1.25 *(Annualizing means and variances) Suppose we have monthly net returns $R_t = P_t/P_{t-1} - 1$. Estimate the means and the standard deviations on monthly returns, and then (when showing the results) multiply the mean by 12 and the standard deviation by $\sqrt{12}$. To see why, notice that an annual return would be*

$$\begin{aligned} P_t/P_{t-12} - 1 &= (P_t/P_{t-1})(P_{t-1}/P_{t-2}) \ldots (P_{t-11}/P_{t-12}) - 1 \\ &= (R_t + 1)(R_{t-1} + 1) \ldots (R_{t-11} + 1) - 1 \\ &\approx R_t + R_{t-1} + \ldots + R_{t-11}. \end{aligned}$$

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

$$\mathrm{E}(R_t + R_{t-1} + \ldots + R_{t-11}) = 12\, \mathrm{E}\, R_t,$$

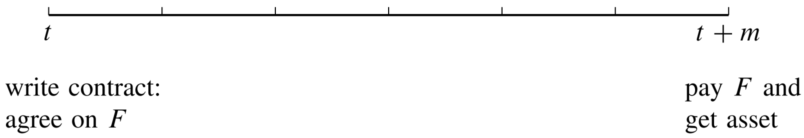


Figure 1.2: Timing convention of forward contract

where $\mathrm{E}\, R_t$ *is the expected monthly return. If returns are iid (in particular, same variance and uncorrelated across time)*

$$\begin{aligned}
\operatorname{Var}(R_t + R_{t-1} + \ldots + R_{t-11}) &= 12 \operatorname{Var}(R_t) \Rightarrow \\
\operatorname{Std}(R_t + R_{t-1} + \ldots + R_{t-11}) &= \sqrt{12} \operatorname{Std}(R_t),
\end{aligned}$$

where $\operatorname{Var}(R_t)$ *is the variance of monthly returns. Instead, with weekly data multiply by 52 and with daily data by 252 (the approximate number of trading days per year). See Table 1.1 for an empirical illustration.*

	Small growth	Small value	Large growth	Large value	Bonds	T-bills	Equity market
mean	0.84	1.19	1.07	1.02	0.52	0.27	0.99
mean (ann.)	10.11	14.32	12.88	12.28	6.29	3.24	11.91
std	6.54	5.33	4.47	4.94	1.34	0.21	4.37
std (ann.)	22.65	18.46	15.50	17.11	4.64	0.73	15.12

Table 1.1: Means and std of asset class returns, US, monthly returns (%), 1985:01-2019:12

1.2 Forward Contracts

A forward contract specifies (among other things) which asset should be delivered at expiration and how much should be paid for it: the forward price, F. See Figure 1.2.

It can be shown that the forward price, F, contracted in t (but to be paid in $t+m$) on an asset without dividends (until the expiration of the forward contract) equals the spot price times an interest rate factor

$$F = S\,(1+Y)^m\,, \tag{1.27}$$

where S is the spot price today (when the forward contract is written) and Y is the interest rate on an m-period loan.

Example 1.26 *With* $Y = 0.05$*,* $m = 0.75$ *and* $S = 100$ *we have the forward price* $F = (1 + 0.05)^{0.75}100 \approx 103.73$.

1.3 Asset Value as Discounted Cash Flow*

1.3.1 Fundamental Asset Value

A *present value* is a sum of discounted future cash flows. A higher discount rate and longer time until the cash flow reduces the present value.

Remark 1.27 *If the cash flow is* -150 *in* t*, 100 in* $t+1$ *and 130 in* $t+2$*, and the discount rate* $R = 0.1$ *then*

$$-150 + \frac{100}{1+R} + \frac{130}{(1+R)^2} \approx 48.3 \textit{ for } R = 0.1.$$

Many assets are long-lived. A fundamental valuation of the asset is that its (fair) price equals the present value of future cash flow

$$\begin{aligned} P_t &= \frac{\mathrm{E}_t\, D_{t+1}}{1+R} + \frac{\mathrm{E}_t\, D_{t+2}}{(1+R)^2} + \frac{\mathrm{E}_t\, D_{t+3}}{(1+R)^3} + \dots \\ &= \textstyle\sum_{s=1}^{\infty} \frac{\mathrm{E}_t\, D_{t+s}}{(1+R)^s}, \end{aligned} \tag{1.28}$$

where D_{t+s} are the future cash flows to the investor. For shares the cash flows are the dividend payments, while for bonds they are the coupon (and face value) payments. In this section, the discount rate R is given (and assumed to be constant). In general, the discount rate depends on both the riskfree rate and the risk of the asset. (This is one of the main topics of the rest of these notes, see, for instance, the discussion of CAPM.) In project evaluation, the discount rate is often a weighted average ("WACC") of the required return on equity and the after tax borrowing rate.

Remark 1.28 *(What if the company cancels dividends in order to invest more?**) Suppose the investment project generates an annual return of ROE—and all earnings are paid out* *in period 3:*

$$\textit{Old plan: } P_0 = \frac{\mathrm{E}_0\, D_1}{1+R} + \frac{\mathrm{E}_0\, D_2}{(1+R)^2} + \frac{\mathrm{E}_0\, D_3}{(1+R)^3} + \dots$$

$$\textit{New plan: } \tilde{P}_0 = \frac{\mathbf{0}}{1+R} + \frac{\mathrm{E}_0\, D_2}{(1+R)^2} + \frac{\mathrm{E}_0\, D_3 + \mathrm{E}_0\, D_1(1+ROE)^2}{(1+R)^3} + \dots$$

Same value ($\tilde{P}_0 = P_0$) if $ROE = R$.

In general, *dividends* reduce the stock price on the ex-dividend day (when the next dividend belongs to the seller, rather than the buyer of the stock) by an amount equal to the dividend. In contrast, a *stock repurchase* does not directly affect the stock price, but clearly reduces the number of outstanding (floating) shares. Both methods (if of same size) are likely to reduce the market value of the firm with the same amount.

Remark 1.29 *(Dividends and stock repurchases*) Suppose the total value of a firm is 100, of which 90 is the present value of future earnings and 10 is cash. With 10 outstanding shares, the share price is 10 (*$100/10$*). If the firm distributes the cash as dividends, then the remaining total value of the firm is 90 so the share price is now 9. Overall the share holders have this period (assuming no other news) received a zero return (dividend yield + capital gain). Instead, if the firm buys back one share at the price of 10, then the total firm value becomes 90 and there are now 9 outstanding shares, so the share price would be unchanged at 10. Again, the return is zero. (Taxes and behavioural aspects can complicate this story.)*

Remark 1.30 *(Valuation in terms of earnings instead of dividends*) Earnings can be spent on dividends or kept on the balance sheet as cash or some other asset (an "investment"): $E = D + I$. The firm value is*

$$P_0 = \frac{\overbrace{\mathrm{E}_0\, D_1}^{E_1 - I_1}}{1+R} + \frac{\overbrace{\mathrm{E}_0\, D_2}^{E_2 - I_2}}{(1+R)^2} + \frac{\overbrace{\mathrm{E}_0\, D_3}^{E_3 - I_3}}{(1+R)^3} + \dots$$

This shows that the firm value equals the present value of future earnings minus the present value of new investment expenditures used to generate those earnings.

Remark 1.31 *(From income to cash flow*) To calculate the cash flow start with the net income (profit) before interests and taxes (EBIT) from the income statement, subtract the taxes (they are costs...), add back the depreciations (it is just an accounting item), subtract* *the capital expenditure (buying machines takes cash, even if it is not booked as a cost) and also subtract the change in the net working capital (current assets minus current liabilities, booked as income but you have not received it yet). All financial transactions are disregarded, so the cash flow must be used to pay all bond and equity holders.*

Remark 1.32 *(Internal Rate of Return) The IRR is the R that makes the net present value of a cash flow process zero. For instance, if the cash flow is -150 in t (an investment), 100 in $t+1$ and 130 in $t+2$, then*

$$-150 + \frac{100}{1+R} + \frac{130}{(1+R)^2} \approx 0 \text{ for } R = 0.32.$$

Typically we have to solve for the IRR by numerical methods. Notice that there may be more than one IRR if the cash flow process changes sign more than once.

1.3.2 "Speculative" Valuation

An alternative view of the asset value is the present of the next dividend plus what you expect to resell the asset for

$$P_t = \frac{\mathrm{E}_t\, D_{t+1} + \mathrm{E}_t\, P_{t+1}}{1+R}. \tag{1.29}$$

This is the same as the fundamental valuation (1.28) *if* you expect to resell it at your (expected next period) fundamental valuation. Otherwise not.

Proof. (of fundamental = speculative asset value, if $\mathrm{E}_t\, P_{t+1}$ follows fundamental valuation) Use (1.28) to write

$$P_{t+1} = \frac{\mathrm{E}_{t+1}\, D_{t+2}}{1+R} + \frac{\mathrm{E}_{t+1}\, D_{t+3}}{(1+R)^2} + \dots.$$

Take expectations as of period t and use in (1.29)

$$P_t = \frac{\mathrm{E}_t\, D_{t+1}}{1+R} + \frac{\mathrm{E}_t\, \mathrm{E}_{t+1}\, D_{t+2}}{(1+R)^2} + \frac{\mathrm{E}_t\, \mathrm{E}_{t+1}\, D_{t+3}}{(1+R)^3} + \dots.$$

Recall that $\mathrm{E}_t(\mathrm{E}_{t+1}\, D_{t+s})) = \mathrm{E}_t\, D_{t+s}$ (the "law of iterated expectations.") to complete the proof. ■

Remark 1.33 *(Law of iterated expectations) The law of iterated expectations implies that*

$$\mathrm{E}_t(\mathrm{E}_{t+1}\, y_{t+2}) = \mathrm{E}_t\, y_{t+2}$$

To see why, let $y_{t+2} = \mathrm{E}_{t+1}\, y_{t+2} + \varepsilon_{t+2}$, so ε_{t+2} is a surprise in $t+2$. The equation above can then be written

$$\mathrm{E}_t(y_{t+2} - \varepsilon_{t+2}) = \mathrm{E}_t\, y_{t+2},$$

which holds if $\mathrm{E}_t\, \varepsilon_{t+2} = 0$. That is, the surprise in $t+2$ cannot be predicted by any information in period t. Basically, this is the same as saying that we know more, not less, as time goes by.

1.3.3 Fundamental Valuation and Returns

The return from holding the asset from t to $t+1$ is

$$R_{t+1} = \frac{D_{t+1} + P_{t+1}}{P_t} - 1. \tag{1.30}$$

If the discount rate in (1.28) is constant over time, then it equals the expected return

$$\mathrm{E}_t\, R_{t+1} = R. \tag{1.31}$$

It follows that if there is no news between t and $t+1$ (so expectations are unchanged, $\mathrm{E}_t\, D_{t+s} = \mathrm{E}_{t+1}\, D_{t+s}$), then

$$R_{t+1} = R \text{ (if no news)}. \tag{1.32}$$

Notice that this return does *not* depend on the level or growth rate of the dividends. Old information is in P_t, and does not affect R_{t+1}.

Proof. (of (1.32)–(1.31)) Use (1.28) to write

$$P_{t+1} = \frac{\mathrm{E}_{t+1}\, D_{t+2}}{1+R} + \frac{\mathrm{E}_{t+1}\, D_{t+3}}{(1+R)^2} + \ldots.$$

Use in the realized return (1.30) and take expectations as of t to get (using $\mathrm{E}_t\, \mathrm{E}_{t+1}\, D_{t+s} = \mathrm{E}_t\, D_{t+s}$)

$$\mathrm{E}_t\, R_{t+1} = \frac{\mathrm{E}_t\, D_{t+1} + \frac{\mathrm{E}_t\, D_{t+2}}{1+R} + \frac{\mathrm{E}_t\, D_{t+3}}{(1+R)^2} + \ldots}{\frac{\mathrm{E}_t\, D_{t+1}}{1+R} + \frac{\mathrm{E}_t\, D_{t+2}}{(1+R)^2} + \frac{\mathrm{E}_t\, D_{t+3}}{(1+R)^3} + \ldots} - 1 = R.$$

In addition, if expectations are unchanged, then $R_{t+1} = \mathrm{E}_t\, R_{t+1}$. (This can also be proved directly by substituting for P_{t+1} in (1.30).) ■

1.3.4 Asset Price with constant Cash Flow Growth

With *constant dividend growth forever* (growing perpetuity), $\mathrm{E}_t\, D_{t+s+1} = (1+g)\,\mathrm{E}_t\, D_{t+s}$, so (1.28) becomes

$$P_t = \mathrm{E}_t\, D_{t+1} \textstyle\sum_{s=1}^{\infty} \frac{(1+g)^{s-1}}{(1+R)^s} = \frac{\mathrm{E}_t\, D_{t+1}}{R-g}. \tag{1.33}$$

This is the "Gordon model." The asset price (1.33) is high when: *(a)* dividends are expected to be high; *(b)* the growth rate (g) is believed to be high; and *(c)* when discounting (R) is low.

Inverting this formula to get the discount rate ("cost of equity capital")

$$R = \frac{\mathrm{E}_t\, D_{t+1}}{P_t} + g. \tag{1.34}$$

Example 1.34 *(Asset price as sum of discounted cash flows) With $D_1 = 100$, $R = 0.1$ and $g = 2\%$,*

$$P_0 = 100/(0.1 - 0.02) = 1250$$

Proof. (of (1.33)) Write the first equality of (1.33) as $P_t = \frac{\mathrm{E}_t\, D_{t+1}}{1+R} \sum_{s=0}^{\infty} (\frac{1+g}{1+R})^s$. Recall the fact that for a geometric series, $\sum_{s=0}^{\infty} r^s = 1/(1-r)$ if $|r| < 1$. Apply this on $r = (1+g)/(1+R)$, to get that

$$P_t = \frac{\mathrm{E}_t\, D_{t+1}}{1+R} \frac{1}{1-(1+g)/(1+R)} = \frac{\mathrm{E}_t\, D_{t+1}}{R-g}.$$

■

1.3.5 Valuation Multiples

The *price-earnings ratio* (p/e) is

$$\text{“}p/e\text{”} = \frac{P}{e} \tag{1.35}$$

(e is short for earnings per share) If dividends are proportional to earnings, $D = k \times e$ and $e_{t+1} = (1+g)e_t$, then

$$p/e = \frac{P_0}{e_0} = \frac{\overbrace{ke_1}^{D_1}/(R-g)}{e_0} = k\frac{1+g}{R-g}.$$

Example 1.35 *$R = 0.1$, $g = 2\%$ and $k = 1$ ("cash cow")*

$$p/e = 1 \times \frac{1.02}{0.1 - 0.02} = 12.75$$

Instead, with $g = 5\%$ we have $p/e = 21$. This shows that p/e is very sensitive to assumptions about the growth rate.

The *multiples approach* is to use a comparison with a peer group (in the market or recent M&A transactions) in order price an asset (here denoted i). It has the advantage that we do not need to specify growth or discount rate. The *equity value method* is calculate the share value of company i as

$$P_i = \left(\frac{P}{e}\right)_{peers} \times e_i, \text{ so } \frac{P_i}{e_i} = \left(\frac{P}{e}\right)_{peers}. \tag{1.36}$$

As alternatives to e, use cash flow and book value. In general, this approach makes sense if firm i and the peers have similar growth and risk, while the dividends might differ.

Remark 1.36 *(The discounted cash flow model vs. the multiples approach) To simplify, assume $D = e$ and assume constant growth. This means that $P = (1+g)e/(R-g)$ for both i and peers. To have $P_i/e_i = (P/e)_{peers}$ as in (1.36), the following must hold*

$$\frac{P_i}{e_i} = \left(\frac{1+g}{R-g}\right)_i = \left(\frac{1+g}{R-g}\right)_{peers} = \left(\frac{P}{e}\right)_{peers}.$$

This shows that the discount and growth rates must be similar.

1.4 Markets, Instruments and Some Key Terms

The initial issuance of an asset (for instance, an IPO) takes place at the *primary market* while the subsequent trading takes place on the *secondary market*. Trading on the secondary market can be done on an *exchange* (NYSE, Tokyo, EuroNext, Nasdaq, London, Shanghai, HK,...,...CBOE, CME,...), an *electronic platform* (EBS, Reuters), or *over the counter* (OTC).

Different asset classes are often traded on different exchanges/platforms. This motives terms like the “money market”, “bond market”, “currency market”, “stock market”, and “derivative markets”.

Remark 1.37 *(On financial instruments)*

- *ETFs are liquid and have low fees*
- *Synthetic vs. physical ETFs (issue: counter party risk)*
- *Futures (more generally, derivatives) often have low transaction costs*

- *Pay high fees only for excellent performance: avoid "closet indexers"*

Remark 1.38 *(On alternative asset classes)*

- *Examples: hedge funds, infrastructure, PE, CoCo, cat bonds*
- *Particularly useful if you are a long-run investor (liquidity...) and willing to pay high fees (pension plans, endowments)*
- *UCITS funds: regulated European funds (can be bought by a broader class of investors). Even Blackstone entered this market 2014 (to sell hedge funds to pension funds)*

Remark 1.39 *(Useful terms) The following stock market terms are useful*

- Market capitalization*: value of all shares*
- Float*: number of not closely held shares*
- Volume*: number of traded shares*
- Short interest*: number of shortened shares*
- Consensus estimate*: the average forecast (of eg. earnings) across analysts*
- ROE*: net income/book value of equity*
- ROI*: (net income + interest rates)/book value of (equity + debt)*

Remark 1.40 *(Trading costs*) As a an investor you typically pay a* commission *(eg. \$25 or \$0.025 per share, whichever is greater) to the broker. In addition, the price depends on whether you are buying (high price) or selling (low price).* Bid and ask *prices are:*

	Definition	*Example*
Ask price	*lowest price at which someone will sell*	90.05
Bid price	*highest price at which someone will buy*	90.00
Bid-ask spread		0.05

If you want to buy immediately*: you submit a* market buy order *(buy at best available price) and you need to pay ask price (90.05). Instead, if you want to* sell immediately, *you submit a* market sell order *and get the bid price (90.00). A round-trip (first buy, then sell) costs* $90.05 - 90.00 = 0.05$ *(the bid-ask spread). Alternatively, you can (at least on some markets) submit a* limit buy order *at a higher bid price (eg. 90.01) or a* limit sell order *at a lower ask price (eg. 90.04). With some luck someone hits that order.*

1.5 Asset Classes

Table 1.2 illustrates the return distributions for different U.S. asset classes look like over a decade. There are distinct differences between small and large firms (the former have higher, but more volatile returns) and between growth and value firms (the latter typically have higher returns). However, the most pronounced difference is between equity and bonds (the latter have much less volatility and often lower returns).

	Small growth	Small value	Large growth	Large value	Bonds	T-bills	Equity market
mean	0.84	1.19	1.07	1.02	0.52	0.27	0.99
std	6.54	5.33	4.47	4.94	1.34	0.21	4.37
min	−32.37	−27.72	−23.18	−22.27	−4.39	0.00	−22.64
max	27.09	17.29	14.44	17.66	5.31	0.79	12.89
corr with market	0.86	0.83	0.97	0.87	−0.07	0.03	1.00
beta against market	1.28	1.02	1.00	0.98	−0.02	0.00	1.00

Table 1.2: Descriptive statistics of asset classes, US, monthly returns (%), 1985:01-2019:12. The beta is the slope coefficient from regressing the asset on the market return.

Figures 1.3–1.4 illustrate the dynamics behind the figures for the entire sample in Table 1.2.

Table 1.3 gives the annual ranking of the asset classes (for a shorter sample). Much of 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.

1.6 Appendix: A Primer in Matrix Algebra*

For this appendix, let c be a scalar and define the matrices

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

Multiplying a matrix by a scalar means multiplying each element by the scalar

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

	6th		5th		4th		3rd		2nd		1st	
2000	SG	-24.8	LG	-13.4	TB	5.9	B	13.5	LV	16.9	SV	24.3
2001	LG	-14.7	SG	0.8	LV	1.5	TB	3.8	B	6.7	SV	23.7
2002	SG	-31.9	LV	-31.0	LG	-23.2	SV	-9.1	TB	1.6	B	11.8
2003	TB	1.0	B	2.2	LG	27.0	LV	27.9	SG	54.1	SV	63.9
2004	TB	1.2	B	3.5	LG	8.2	SG	15.4	LV	19.7	SV	20.2
2005	SG	-0.0	B	2.8	TB	3.0	LG	4.5	SV	9.0	LV	12.1
2006	B	3.1	TB	4.8	SG	8.8	LG	11.0	LV	23.6	SV	24.5
2007	SV	-11.1	LV	-0.0	TB	4.7	SG	5.5	B	9.0	LG	12.6
2008	SG	-40.2	LV	-39.6	LG	-33.8	SV	-32.9	TB	1.6	B	13.7
2009	B	-3.6	TB	0.1	LV	19.2	SV	31.1	LG	31.7	SG	37.0
2010	TB	0.1	B	5.9	LV	7.9	LG	15.5	SV	26.6	SG	29.6
2011	LV	-10.1	SV	-8.5	SG	-5.7	TB	0.0	LG	3.8	B	9.8
2012	TB	0.1	B	2.0	SG	14.9	LG	15.0	SV	19.9	LV	29.4
2013	B	-2.7	TB	0.0	LG	33.9	LV	40.1	SV	42.0	SG	45.2
2014	TB	0.0	SV	3.9	B	5.1	SG	5.3	LV	11.8	LG	13.8
2015	SV	-9.7	LV	-7.8	SG	-3.0	TB	0.0	B	0.8	LG	4.8
2016	TB	0.2	B	1.0	SG	7.9	LG	8.8	LV	25.8	SV	36.6
2017	TB	0.8	B	2.3	SV	9.6	LV	18.3	SG	25.6	LG	30.2
2018	LV	-14.1	SV	-12.6	SG	-8.2	LG	-0.0	B	0.9	TB	1.8
2019	TB	2.1	B	6.9	SV	14.1	LV	26.4	SG	30.6	LG	34.2

Table 1.3: 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.

Example 1.41 *(Matrix × scalar)*

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

Adding/subtracting a scalar to each element of a matrix can be done by

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

where J is a matrix (of the same size as A) filled with ones. This is sometimes written $A + c$, although that notation is not universally liked. In some applications, $\mathbf{1}_n$ (or just $\mathbf{1}$) is used to denote a vector of n ones.

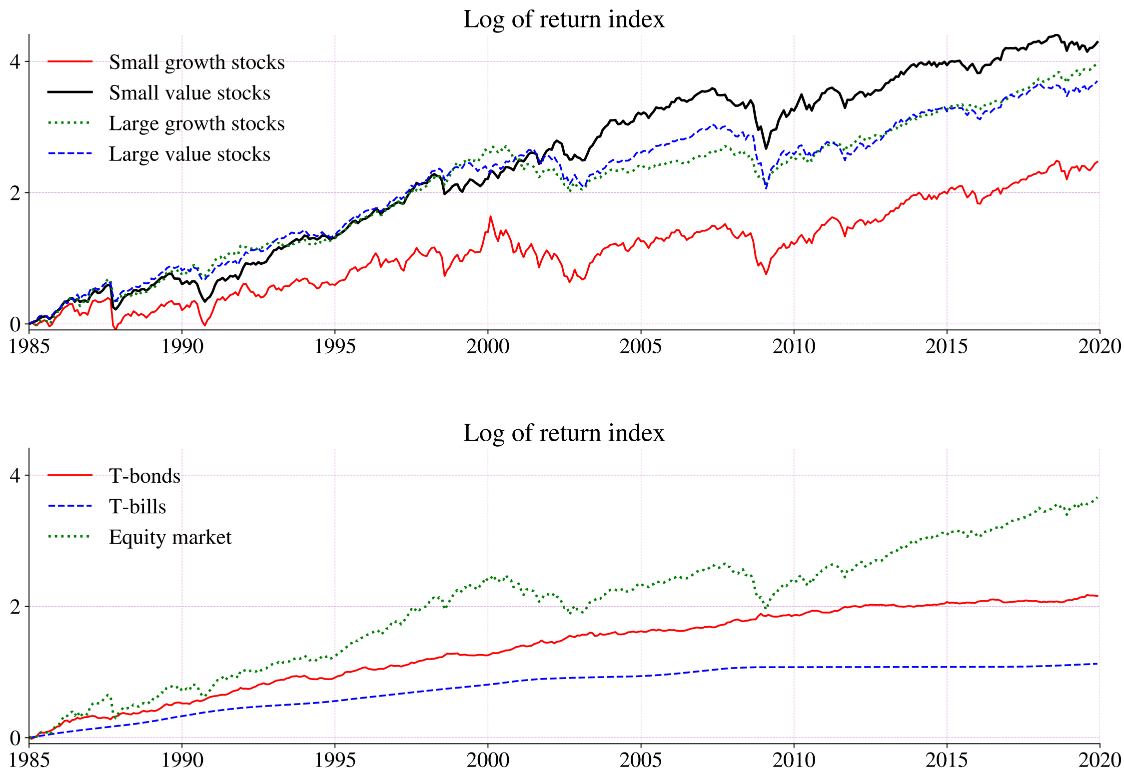


Figure 1.3: Performance of US equity and fixed income

Example 1.42 *(Matrix $\pm$ scalar)*

$$\begin{bmatrix} 10 \\ 11 \end{bmatrix} - 10 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

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

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

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

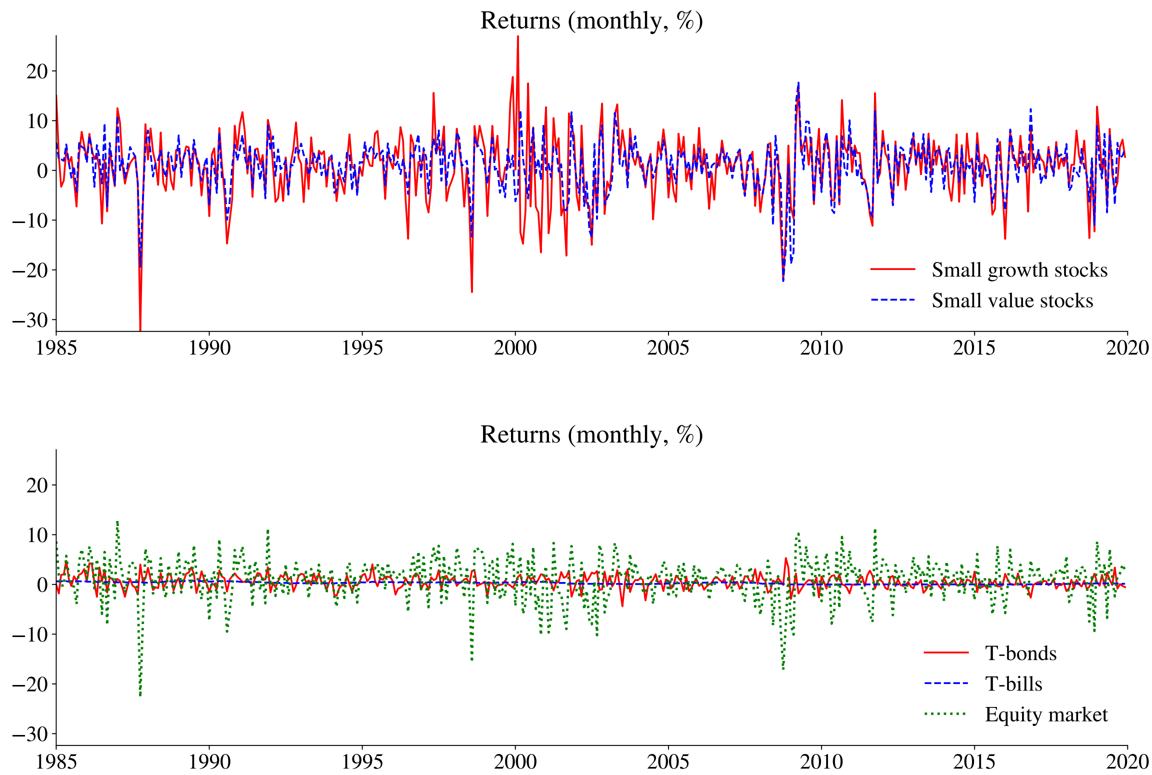


Figure 1.4: Performance of US equity and fixed income

Example 1.43 *(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 1.44 *(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 1.45 *(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* (a scalar)

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

and xz' the *outer product* (a matrix)

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

(Notice that xz does not work).

Example 1.46 *(Inner and outer products)*

$$\begin{bmatrix}10\\11\end{bmatrix}'\begin{bmatrix}2\\5\end{bmatrix} = \begin{bmatrix}10 & 11\end{bmatrix}\begin{bmatrix}2\\5\end{bmatrix} = 75$$
$$\begin{bmatrix}10\\11\end{bmatrix}\begin{bmatrix}2\\5\end{bmatrix}' = \begin{bmatrix}10\\11\end{bmatrix}\begin{bmatrix}2 & 5\end{bmatrix} = \begin{bmatrix}20 & 50\\22 & 55\end{bmatrix}$$

If x is a column vector and A a square matrix, then the product $x'Ax$ is a quadratic form *(a scalar).*

Example 1.47 *(Quadratic form)*

$$\begin{bmatrix}10\\11\end{bmatrix}'\begin{bmatrix}1 & 3\\3 & 4\end{bmatrix}\begin{bmatrix}10\\11\end{bmatrix} = 1244$$

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 zeros elsewhere). The matrix inverse is useful for solving systems of linear equations, $y = Ax$ as $x = A^{-1}y$.

Example 1.48 *(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 1.49 *(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 1.50 *(Derivative of a quadratic form) With $n = 2$, the quadratic form is*

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

The derivatives with respect to x_1 and x_2 are

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

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

1.7 Appendix: Data Sources*

The data used in these lecture notes are from the following sources:

1. website of Kenneth French,
   `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`
2. Datastream
3. Federal Reserve Bank of St. Louis (FRED), `http://research.stlouisfed.org/fred2/`
4. website of Robert Shiller, `http://www.econ.yale.edu/~shiller/data.htm`
5. yahoo! finance, `http://finance.yahoo.com/`
6. OlsenData, `http://www.olsendata.com`

1.7 Appendix: A Primer in Matrix Algebra*

1.7.1 Adding and Multiplying: A Matrix and a Scalar

For this appendix, let c be a scalar and define the matrices

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

Multiplying a matrix by a scalar means multiplying each element by the scalar

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

Example 1.51 *(Matrix × scalar)*

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

Adding/subtracting a scalar to each element of a matrix can be done by

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

where J is a matrix (of the same size as A) filled with ones. This is sometimes written $A + c$, although that notation is not universally liked. In some applications, $\mathbf{1}_n$ (or just $\mathbf{1}$) is used to denote a vector of n ones.

Example 1.52 *(Matrix ± scalar)*

$$\begin{bmatrix} 10 \\ 11 \end{bmatrix} - 10 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix} + 10 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 13 \\ 13 & 14 \end{bmatrix}.$$

1.7.2 Adding and Multiplying: Two Matrices

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 1.53 *(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'

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

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 1.54 *(Matrix multiplication)*

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

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

1.7.3 Transpose

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 1.55 *(Matrix transpose)*

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

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

1.7.4 Inner and Outer Products, Quadratic Forms

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

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

and xz' the *outer product* (a matrix)

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

(Notice that xz does not work).

Example 1.56 *(Inner and outer products)*

$$\begin{bmatrix} 10 \\ 11 \end{bmatrix}' \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 10 & 11 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \end{bmatrix} = 75$$

$$\begin{bmatrix} 10 \\ 11 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \end{bmatrix}' = \begin{bmatrix} 10 \\ 11 \end{bmatrix} \begin{bmatrix} 2 & 5 \end{bmatrix} = \begin{bmatrix} 20 & 50 \\ 22 & 55 \end{bmatrix}$$

If x is a column vector and A a square matrix, then the product $x'Ax$ is a quadratic form *(a scalar).*

Example 1.57 *(Quadratic form)*

$$\begin{bmatrix} 10 \\ 11 \end{bmatrix}' \begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 10 \\ 11 \end{bmatrix} = 1244$$

1.7.5 Matrix Inverse

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 zeros elsewhere). The matrix inverse is useful for solving systems of linear equations, $y = Ax$ as $x = A^{-1}y$.

Example 1.58 *(Matrix inverse) We have*

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

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

1.7.6 OLS Notation: $X'X$ or $\sum_{t=1}^{T} x_t x_t'$?

Let x_t be a $K \times 1$ vector of (of data in period t). We can the calculate the outer product ($K \times K$) as $x_t x_t'$ and summing each element across T observations gives the $K \times K$ matrix $S_{xx} = \sum_{t=1}^{T} x_t x_t'$.

Alterntively, let X be a $T \times K$ matrix with x_t' in row t. Then we can also calculate S_{xx} as $X'X$.

Example 1.59 *(Sum of outer product, $\sum_{t=1}^{T} x_t x_t'$)*

$$x_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, x_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \text{and } x_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

We then have

$$\begin{aligned} \sum_{t=1}^{T} x_t x_t' &= \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}. \end{aligned}$$

In this example, the matrix happens to be diagonal, but that is not a general result. However, it will always be symmetric.

Example 1.60 *(Sum of outer product, $X'X$) Define*

$$X = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

It is straightforward to calculate that $X'X = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$.

1.7.7 Derivatives of Matrix Expressions

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

Example 1.61 *(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 1.62 *(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*

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

Chapter 2

The Basics of Portfolio Choice

There are two key elements to portfolio choice: (1) how to mix various risky assets to create a well diversified portfolio; and (2) how to mix the risky assets (leverage) with a riskfree asset to handle the overall risk level.

2.1 Investor Preferences

We never know what the returns on our portfolio will be, but we can formulate beliefs—and base the portfolio choice on those. Perhaps surprisingly, historical return distributions are fairly good predictors of the future return patterns. (In practice, judgemental adjustments are often made to the historical evidence to produce the required inputs to the portfolio choice rules.)

We will (initially) let the investor pick a portfolio that strikes a balance between high expected returns and the variance (as a proxy for risk). For a symmetric returns distribution, the variance is proportional to the downside risk (which we may care more about). For instance, see Figure 2.1 for an illustration of how volatility relates to risk. For instance, suppose you had to (because of liquidity reasons) to sell off the portfolio in 2008–2009. Clearly some assets are riskier than other assets.

2.2 Leverage

2.2.1 A Portfolio of a Single Risky Asset and a Riskfree Asset

Suppose you can only invest into a risky asset (with return R_i) and a riskfree (with return R_f). You may think of the risky asset as being the (equity) market portfolio. To see the

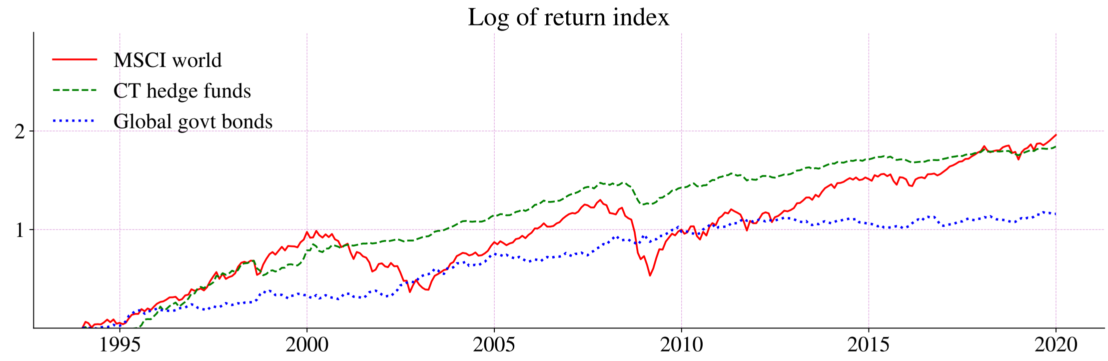


Figure 2.1: Return indices for several asset classes

effect on the mean and the volatility of the portfolio choice, notice that

$$R_p = vR_i + (1-v)R_f \text{, so} \tag{2.1}$$

$$\mathrm{E}\, R_p = v\mu_i + (1-v)R_f \text{ and} \tag{2.2}$$

$$\mathrm{Std}(R_p) = |v|\sigma_i. \tag{2.3}$$

The result on the standard deviation follows from $\mathrm{Var}(R_p) = v^2\sigma_{ii}$, since $\mathrm{Var}(R_f) = 0$ (the riskfree rate over the investment horizon is known when the portfolio is formed) and hence also the covariance is zero. If we use an interest rate to represent the riskfree rate, then we should typically use a maturity that corresponds to the investment horizon.

How much to put into the risky asset is a matter of *leverage*, and v is often called the *leverage ratio*. We typically define the leverage ratio as the investment (into risky assets) divided by how much capital we own, that is, as the portfolio weight v in (2.1).

Example 2.1 *(Leveraged portfolios) Portfolio weights for three different portfolios*

	Portfolio A	*Portfolio B*	*Portfolio C*	*Portfolio D*
v *(in risky assets)*	0.5	1	2	-1
$1-v$ *(in riskfree)*	0.5	0	-1	2
Sum	1	1	1	1

Portfolio A*: your capital is* 200*, invest* 100 *in risky assets and* 100 *in riskfree;* Portfolio B*: your capital is* 200*, invest* 200 *in risky assets and* 0 *in riskfree;* Portfolio C*: your capital is* 200*, invest* 400 *in risky assets and* -200 *in riskfree (borrow 200 = short position in* *riskfree).*

Example 2.2 *(Short-selling) We also consider a portfolio D which has* $v = -1$ *and* $1 - v = 2$*. This means that we short-sell the risky asset and put all the money on a bank account.*

Remark 2.3 *(Assuming that* R_i *and* R_f *do not depend on* v*) These notes (mostly) assume that the portfolio choice (here* v*) does not affect the returns. This means that we assume that the investor is small compared to the overall market. It also means that we effectively assume that lending* $(1 - v > 0)$ *and borrowing* $(1 - v < 0)$ *can be done at the same rate. This is a reasonable approximation for a large financial institution and simplifies the analysis considerably.*

Both the mean and the standard deviation are scaled by the leverage ratio v. Notice that taking on leverage (borrowing to invest into the risky asset) typically is a way to increase the expected return of the portfolio, but at the cost of increasing the risk. In terms of the return distribution that amounts to shifting the mean up, but also to making the distribution wider (and thus increasing the probability of a really bad outcome). Figure 2.2 provides an empirical example and Figure 2.3 illustrates the effect on the portfolio return distribution.

Example 2.4 *With* $R_i \sim N(0.095, 0.08^2)$ *(mean of 9.5%, standard deviation of 8%) and* $R_f = 0.03$*, we get (in %)*

	Portfolio A	*Portfolio B*	*Portfolio C*
Mean	6.25	9.5	16
Std	4	8	16

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

$$\mathrm{E}\, R_p = R_f + \mathrm{Std}(R_p) \times SR_i, \tag{2.4}$$

where $SR_i = \left(\mathrm{E}\, R_i - R_f\right) / \mathrm{Std}(R_i)$ is the *Sharpe ratio* of the risky asset. This shows that the average portfolio return is linearly related to its standard deviation. See Figure 2.3.

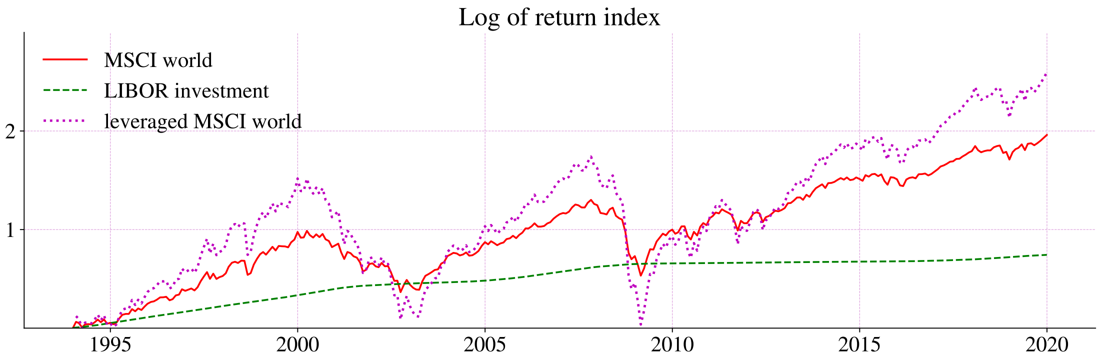


Figure 2.2: The effect of leverage on the portfolio performance

2.2.2 Optimal Portfolio of a Single Risky Asset and a Riskfree Asset

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 one riskfree asset (T-bill, say), the investor maximizes expected utility, denoted $\mathrm{E}\, U(R_p)$, by chosing the weight on the risky asset (the remainder going to the riskfree asset). That is, the investor maximizes

$$\mathrm{E}\, U(R_p) = \mathrm{E}\, R_p - \frac{k}{2}\operatorname{Var}(R_p), \text{ where} \tag{2.5}$$

$$\begin{aligned} R_p &= vR_i + (1-v)R_f \\ &= vR_i^e + R_f. \end{aligned} \tag{2.6}$$

The parameter 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)

$$\begin{aligned} \mathrm{E}\, U(R_p) &= \mathrm{E}(vR_i^e + R_f) - \frac{k}{2}\operatorname{Var}(vR_i^e + R_f) \\ &= v\mu_i^e + R_f - \frac{k}{2}v^2\sigma_{ii}, \end{aligned} \tag{2.7}$$

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

Remark 2.5 *(First order condition for optimising a differentiable function). We want to find the value of b in the interval $b_{low} \leq b \leq b_{high}$, which makes the value of the differentiable function $f(b)$ as large (or small) as possible. The answer is b_{low}, b_{high}, or a value of b where $df(b)/db = 0$.*

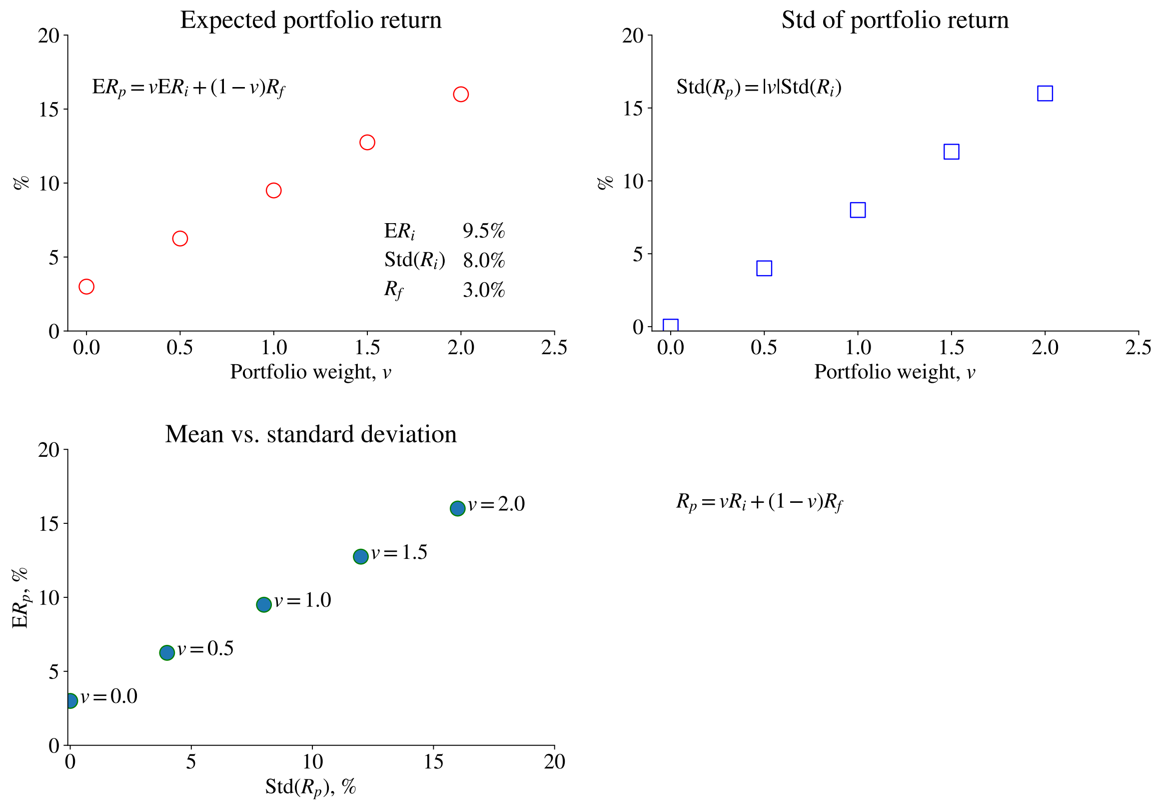


Figure 2.3: The effect of leverage on the portfolio return distribution

The first order condition for an optimum is

$$0 = \partial \mathrm{E}\, U(R_p)/\partial v = \mu_i^e - k v \sigma_{ii}, \tag{2.8}$$

so the optimal portfolio weight of the risky asset is

$$v = \frac{1}{k} \frac{\mu_i^e}{\sigma_{ii}}. \tag{2.9}$$

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 2.6 *(Portfolio choice) If $\mu_i^e = 6.5\%$, $\sigma_i = 8\%$ and $k = 25$, then $v \approx 0.41$. Instead, with $k = 10$, $v \approx 1.02$. See Figure 2.4.*

Remark 2.7 *(Scaling of returns) Returns are often expressed in percents, that is, multiplied by 100. Clearly, this scales up means and standard deviations by 100, so e.g. the*

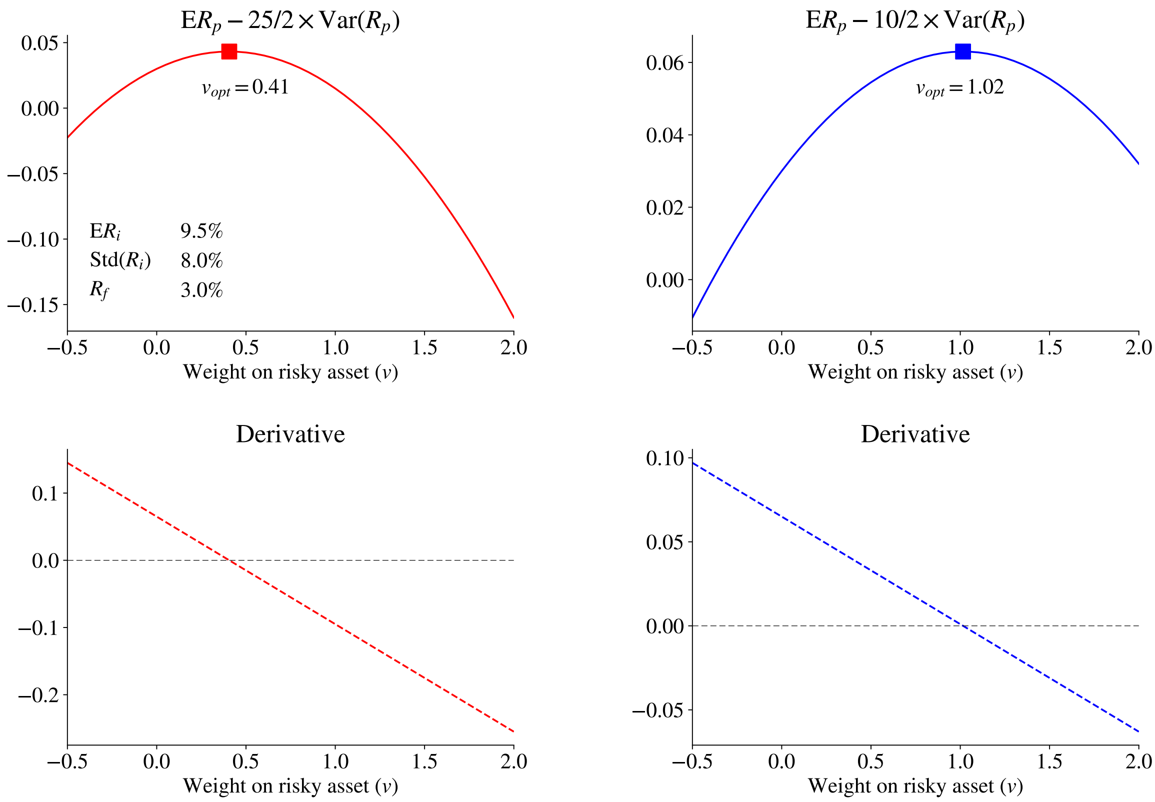


Figure 2.4: Portfolio choice

Sharpe ratio is the same as for the original non-scaled returns. However, this multiplication scales up the variance (and covariance) by 100^2*, so care is needed when using this scaled return in further calculations that involve both means and variances (or covariances). For instance, applied to percentage returns (2.9) would give* $6.5/(25 \times 8^2) \approx 0.004$.

Example 2.8 *The optimal portfolio weight (2.9) is unchanged if both* μ_i^e *and* σ_{ii} *are scaled by (for instance) 100. Notice, however, that if you have return data and multiply it by 100 (to get percentage returns) before estimating the mean and the variance, then that scales the average returns by 100 but the variance by* 100^2*. Using these numbers would clearly change the optimal solution.*

This optimal solution implies that

$$\frac{\mathrm{E}\, R_p^e}{\mathrm{Var}(R_p)} = k, \tag{2.10}$$

where R_p is the portfolio return (2.6) obtained by using the optimal v (from (2.9)). 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 (2.10)) We have

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

which by using (2.9) gives (2.10). ■

Remark 2.9 *(Why not use* $\mathrm{E}\, R_p - k\, \mathrm{Std}(R_p)$*?) Because it is very tricky and there may not be a solution. Too see this, consider changing (2.7) to* $v\mu_i^e + R_f - k\sqrt{v^2\sigma_{ii}}$*. Suppose* $\mu_i^e > 0$ *so we would expect* $v \geq 0$ *to be optimal. Then we can write the expression as* $v(\mu_i^e - k\sigma_i) + R_f$*. If the term in parenthesis is positive, then* $v = \infty$ *is optimal. Instead, if the term is negative, then* $v = 0$ *is optimal. Finally, if the term is zero, then all* $v \geq 0$ *values are equally good. The problem is that both the average returns and the "volatility" are linear in* v*. Instead, if we were to maximize* $v\mu_i^e + R_f - k(v^2\sigma_{ii})^{0.51}$ *(notice the* 0.51 *instead of* 0.5*), then the problem is well behaved.*

2.3 Diversification

This section demonstrates that the portfolio variance can be reduced by (a) mixing assets that are only weakly correlated and (b) mixing many assets in the portfolio. These diversification benefits can often be achieved without hurting the expected returns.

Recall that the variance of a portfolio return is

$$\mathrm{Var}(R_p) = \sum_{i=1}^{n} w_i^2\sigma_{ii} + \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_i w_j \sigma_{ij} = w'\Sigma w, \tag{2.11}$$

where w_i is the portfolio weight on asset i, σ_{ii} is the variance of asset i and σ_{ij} is the covariance of assets i and j. For instance, with 2 (risky) assets we have

$$\mathrm{Var}(R_p) = w_1^2\sigma_{11} + w_2^2\sigma_{22} + 2w_1w_2\sigma_{12}. \tag{2.12}$$

Remark 2.10 *(Covariances and correlations) Recall that* $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$ *or* $\rho_{ij} = \sigma_{ij}/(\sigma_i\sigma_j)$.

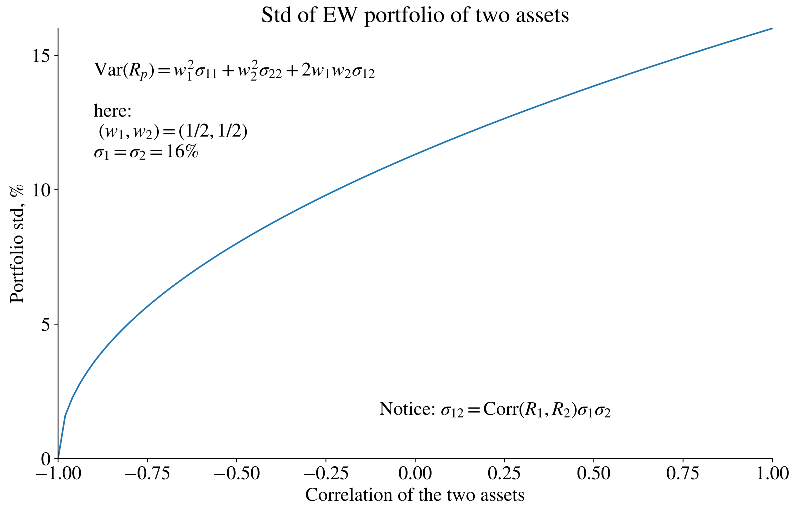


Figure 2.5: Effect of correlation on the diversification benefits

2.3.1 Diversification: The Correlations

As a simple example, consider an *equally weighted (EW) portfolio* of two risky assets (use $w_1 = w_2 = 1/2$ in (2.11)). Denote the correlation by ρ and write as (since $\sigma_{12} = \rho\sigma_1\sigma_2$)

$$\begin{aligned} \text{Var}(R_p) &= \frac{1}{4}\sigma_{11} + \frac{1}{4}\sigma_{22} + \frac{1}{2}\rho\sigma_1\sigma_2 \\ &= \frac{1}{2}\sigma_{11}(1+\rho) \text{ if } \sigma_{11} = \sigma_{22}, \end{aligned} \tag{2.13}$$

where the second equality assumes that both assets have the same variance.

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 even stronger when the correlation becomes negative: with $\rho = -1$ the portfolio variance 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 2.5 for an illustration.

Example 2.11 *(Diversification) If $\sigma_1 = 16\%$ (so $\sigma_{11} = 256/100^2$) and $\rho = 0.5$, then (2.13) gives* $\text{Var}(R_p) = 192/100^2$ *and thus* $\text{Std}(R_p) \approx 13.9\%$.

Table 2.1 gives an empirical examples of the correlations between major asset classes.

	Small growth	Small value	Large growth	Large value	Bonds	T-bills	Equity market
Small growth	1.00	0.86	0.81	0.68	−0.17	−0.03	0.86
Small value	0.86	1.00	0.73	0.83	−0.16	−0.01	0.83
Large growth	0.81	0.73	1.00	0.78	−0.03	0.05	0.97
Large value	0.68	0.83	0.78	1.00	−0.10	0.05	0.87
Bonds	−0.17	−0.16	−0.03	−0.10	1.00	0.20	−0.07
T-bills	−0.03	−0.01	0.05	0.05	0.20	1.00	0.03
Equity market	0.86	0.83	0.97	0.87	−0.07	0.03	1.00

Table 2.1: Correlations of asset class returns, US, monthly returns, 1985:01-2019:12

2.3.2 Diversification: The Number of Assets

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 basic idea of diversification clearly holds also for portfolios that are not equally weighted.

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

$$\mathrm{Var}(R_p) = \sum_{i=1}^{n} \overbrace{\frac{1}{n^2}}^{w_i^2} \sigma_{ii} = \frac{1}{n} \sum_{i=1}^{n} \frac{\sigma_{ii}}{n} \tag{2.14}$$

$$= \frac{1}{n} \overline{\sigma}_{ii}, \text{ (if } \sigma_{ij} = 0). \tag{2.15}$$

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, (2.15) 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

$$\mathrm{Var}(R_p) = \frac{1}{n}(\overline{\sigma}_{ii} - \overline{\sigma}_{ij}) + \overline{\sigma}_{ij}, \tag{2.16}$$

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.

Example 2.12 *(Variance of portfolio return) With* $\overline{\sigma}_{ii} = 256/100^2$ *and* $\overline{\sigma}_{ij} = 128/100^2$, *we get a portfolio variance of* $(256, 192, 170.7)/100^2$ *for* $n = (1, 2, 3)$, *and thus portfolio standard deviations of* $(16\%, 13.9\%, 13.1\%)$.

See Figure 2.6 for an empirical example. Also, Figure 2.7 shows the contributions (according to (2.16)) of the variances and the covariances to the portfolio variance. Clearly, the covariances start to dominate as the number of assets in the portfolio increases—and the portfolio variance goes towards the average covariance. Figure 2.8 suggests that the diversification benefits are not constant across time.

Proof. (of (2.16)) The portfolio variance is

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

which can be rearranged as (2.16). ■

Remark 2.13 *(On negative covariances in (2.16)*) 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 (2.16) this manifests itself in that* $\overline{\sigma}_{ij} < 0$ *would give a negative portfolio variance as n increases.*

2.4 Portfolio Management

Remark 2.14 *(A broad classification of portfolio management strategies)*

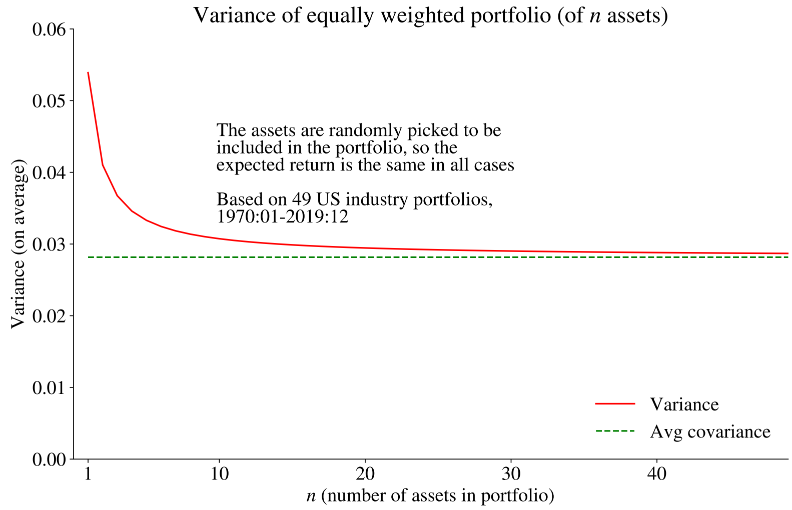


Figure 2.6: Effect of diversification

- *Index management: passive (1975: "Bogle's Folly", 2015: more than 200 billion USD in Vanguard 500)*
- *Fundamental analysis: active, study the economic fundamentals—belief in long-run predictability*
- 1. *top-down (focus on asset classes or industries) or bottom-up (stock-picking)*
  2. *Value (buy cheap firms) or growth (buy promising firms) investing*
     (a) *growth-oriented investors often focus on growth in EPS ("the business story") and expects P/E to remain constant*
     (b) *value-oriented investors often focus on the P/E and expect it to change ("correction" of the market's current valuation of the firm)*
- *Technical analysis: active, study trends—belief in short-run predictability*

Remark 2.15 *(The performance of retail investors) Overall, retail investors seem to do worse than institutional investors. Some possible reasons:*

- *trade too much (trading costs)*
- *pay high fees (the house bank's mutual fund...)*
- *are behind the curve (dotcom fund flows)*
- *get too emotional ("You are your own worst enemy")*

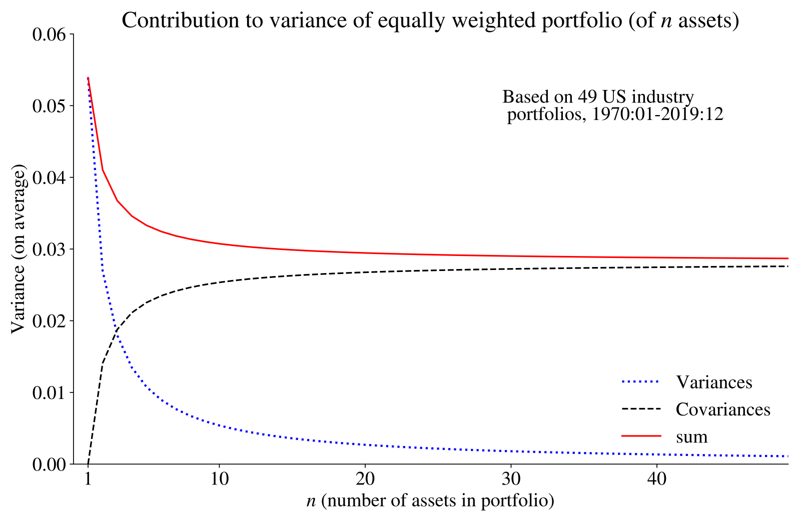


Figure 2.7: Contributions of variances and covariances to the portfolio variance

- *buy complicated (structured) products*
- *diversify too little, stock picking: overconfidence ("know what you don't know")*
- *buy funds that used to outperform the market*
- *do not have access to the very best funds (hedge funds, private equity funds)*

2.5 Appendix: 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 2.9 for the surface of the loss function and the contours.

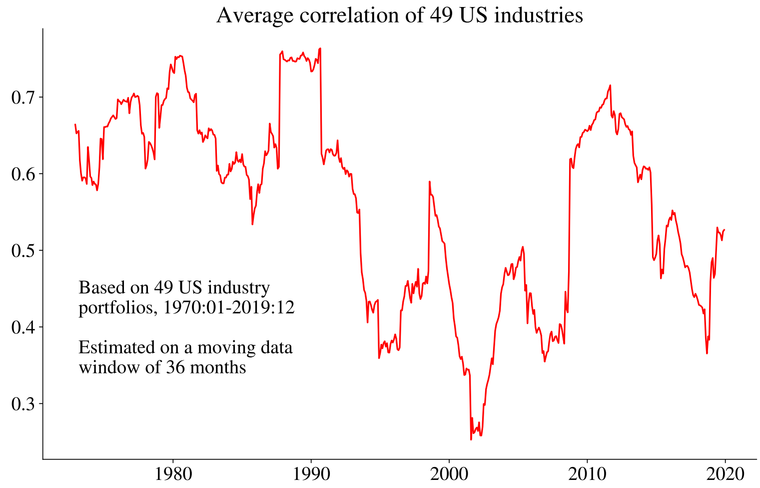


Figure 2.8: Time-varying correlations

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

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

which requires $y = -1/2$, which by implies $x = 4$. (We could equally well have substituted for y). This is also the unique solution. See Figure 2.10. This is an easy way to eliminate an equality restriction.

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

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

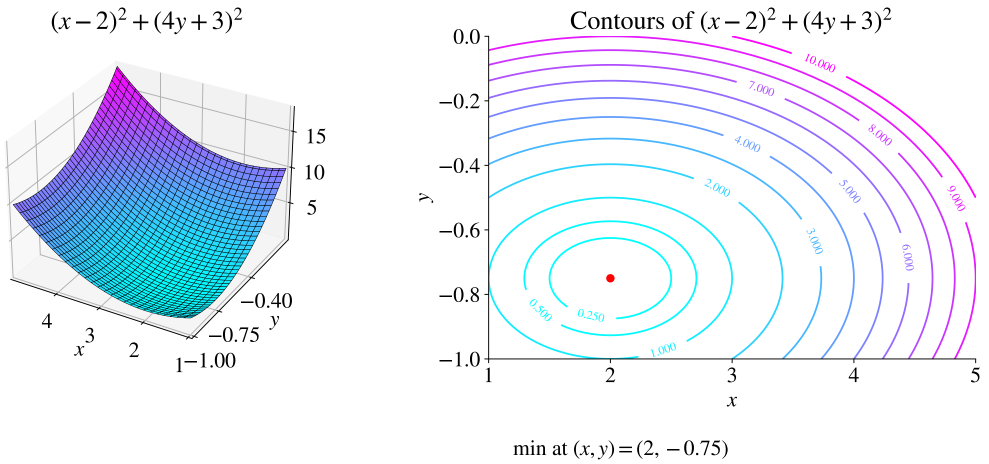


Figure 2.9: Minimization problem

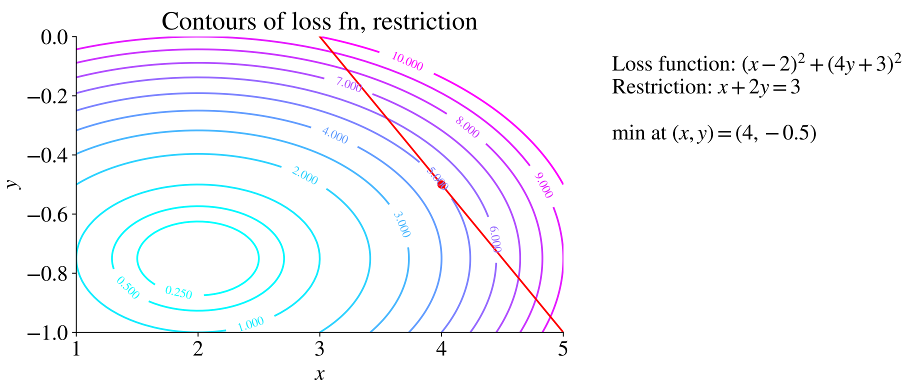


Figure 2.10: Minimization problem with restriction

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

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

These are 3 equations in 3 unknowns (x, y, λ) which we have to solve. One way is as follows. The first two conditions say

$$
\begin{aligned}
x &= \lambda/2 + 2 \\
y &= \lambda/16 - 3/4,
\end{aligned}
$$

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)

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

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 there are several restrictions and/or when the restriction(s) become complicated.

Chapter 3

Mean-Variance Frontier

Reference: Elton, Gruber, Brown, and Goetzmann (2014) 5–6

3.1 Mean-Variance Frontier of Risky Assets

The mean-variance frontier is based on the idea that the investor wants high average portfolio returns, but dislikes portfolio return variance. Although high variances means both large downsides and upsides, the former typically hurts more than the latter pleases—so variance is considered bad. (Later in the course we will look at also other measures of risk.)

	μ, %	Σ, bp		
		A	B	C
A	11.5	166	34	58
B	9.5	34	64	4
C	6.0	58	4	100

Table 3.1: Characteristics of the assets in the MV examples. Notice that μ, % is the expected return in % (that is, $\times 100$) and Σ, bp is the covariance matrix in basis points (that is, $\times 100^2$).

Example 3.1 *(Mean and Std of a portfolio) Table 11.1 illustrates a case with three investable assets (A, B and C). The mean returns are given in %, so 6% should be read as 0.06. In contrast, the covariance matrix is given in terms of basis points (bp, where 1bp $= 1/10000$), so 64bp. should be read as 0.0064. The square root of a variance is the standard deviation, so $\sqrt{0.0064} = 0.08$, that is, 8%. Figure 3.1 also illustrates some portfolios $(1, 2, 3)$ based on the three investable assets.*

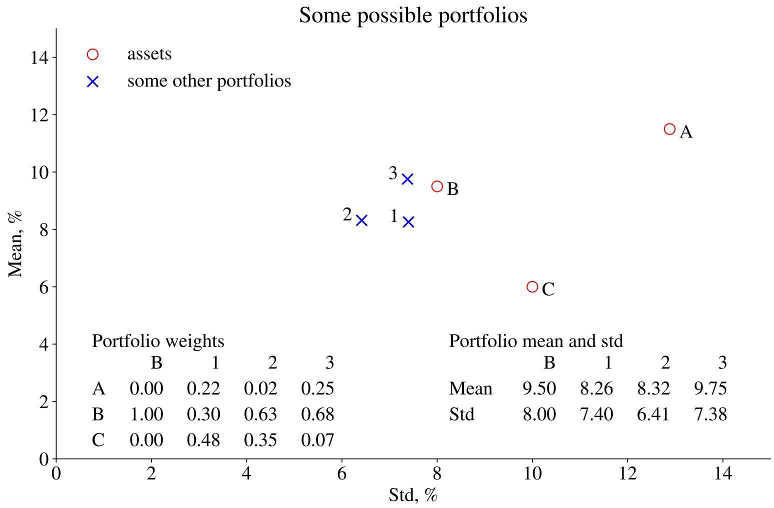


Figure 3.1: Mean vs standard deviation. The properties of the investable assets (A, B, and C) are shown in Table 11.1.

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

$$\begin{gathered}\min_{w_i} \text{Var}(R_p) \text{ subject to} \\ \text{E}\, R_p = \mu^* \text{ and } \textstyle\sum_{i=1}^{n} w_i = 1.\end{gathered} \tag{3.1}$$

Let μ be the $n \times 1$ vector of average returns of all n investable assets, Σ the $n \times n$ covariance matrix of the returns and w the $n \times 1$ vector of portfolio weights.

Remark 3.2 *(Portfolio average and variance) The portfolio mean and variance can be calculated as*

$$\begin{aligned}\text{E}\, R_p &= w'\mu \\ \text{Var}(R_p) &= w'\Sigma w.\end{aligned}$$

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 expected 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

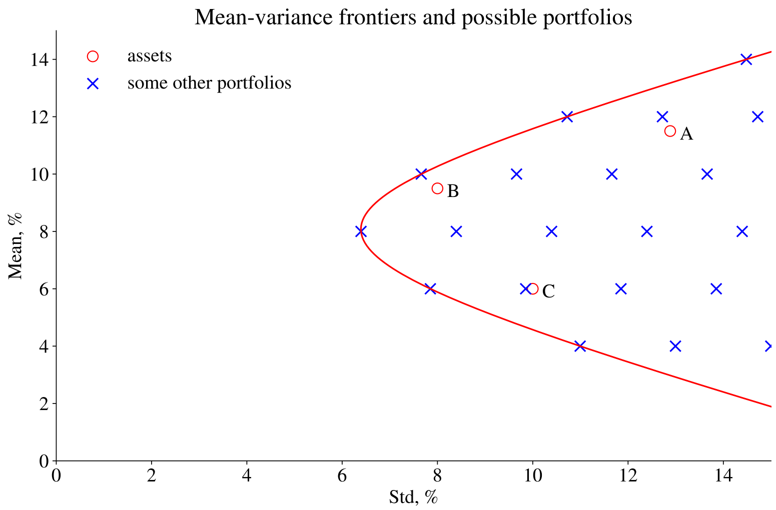


Figure 3.2: Mean-variance frontiers

return). Notice that there are no portfolios (based on the given original assets) that are above or to the left of the efficient frontier. See Figure 3.2 for an example.

Remark 3.3 *(How many different portfolios are there with* $\mathrm{E}\,R_p = \mu^*$*?) With two assets,* $\mathrm{E}\,R_p = w'\mu$ *and* $\Sigma_{i=1}^{n} w_i = 1$ *give* $w\mu_1 + (1-w)\mu_2 = \mu^*$*, and there is only one choice of* w *that satisfy this (assuming* $\mu_1 \neq \mu_2$*). Instead with three assets, we require* $w_1\mu_1 + w_2\mu_2 + (1 - w_1 - w_2)\mu_3 = \mu^*$ *which can hold for a continuum of* (w_1, w_2) *values.*

Remark 3.4 *(Minimizing* $\mathrm{Std}(R_p)$ *instead) We get the same solution if we instead minimize* $\mathrm{Std}(R_p)$ *subject to the same restrictions as in (3.1). This follows from the fact that the square root is an increasing function (and we know that* $\mathrm{Var}(R_p) \geq 0$*). (The sections below give more details.)*

3.1.1 Mean Variance Frontier with Two Risky Assets

In the case of only two investable assets, the MV frontier can be calculated by simply calculating the mean and variance

$$\mathrm{E}\,R_p = w\mu_1 + (1-w)\mu_2 \tag{3.2}$$

$$\mathrm{Var}(R_p) = w^2\sigma_{11} + (1-w)^2\sigma_{22} + 2w(1-w)\sigma_{12}. \tag{3.3}$$

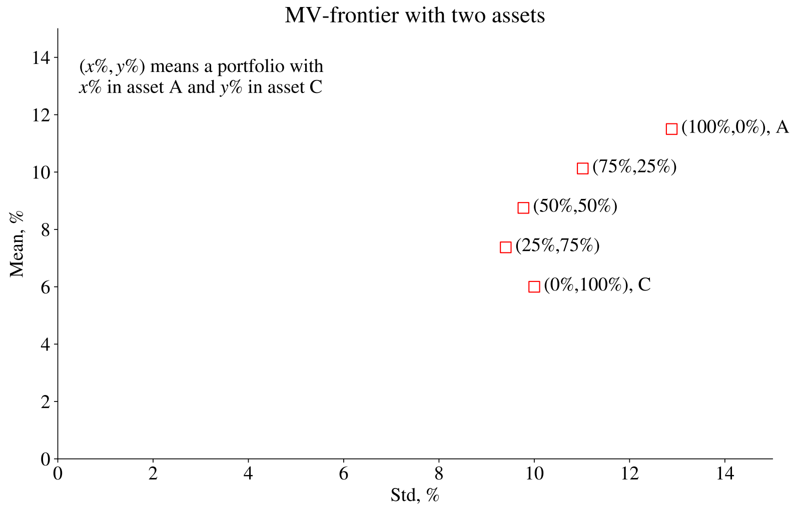


Figure 3.3: Mean-variance frontiers for two risky assets

at a set of different portfolio weights (for instance, $w = (-1, -0.5, 0, 0.5, 1)$.) The reason is that, with only two assets, both are on the MV frontier—so no explicit minimization is needed. See Figure 3.3 for an example.

Remark 3.5 *(Why are both assets on the MV frontier?) Short answer: because there is only one choice of* w *that makes* $\mathrm{E}\,R_p$ *in (3.2) equal to the required expected return* μ^*.

3.1.2 Mean Variance Frontier with Portfolio Restrictions

It is (relatively) straightforward to calculate the mean-variance frontier if there are no other constraints: it just takes some linear algebra (the details are found in later sections).

There are sometimes *additional restrictions*, for instance,

$$\text{no short sales: } w_i \geq 0. \tag{3.4}$$

We then have to apply some explicit numerical minimization algorithm to find portfolio weights. See Figure 3.4 for an example. Algorithms that solve quadratic problems are best suited.

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

$$|w_i^{new} - w_i^{old}| < U_i, \tag{3.5}$$

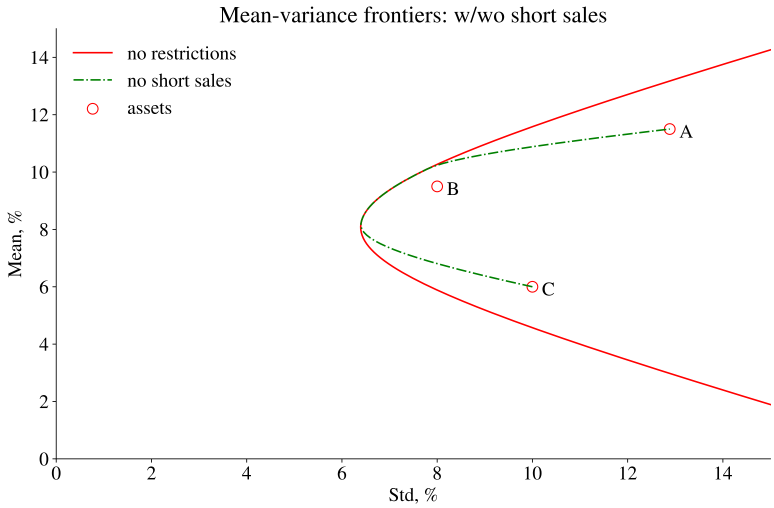


Figure 3.4: Mean-variance frontiers

or that the portfolio weights must be between some boundaries

$$L_i \leq w_i \leq U_i. \tag{3.6}$$

For instance, many mutual funds cannot put more than 10% of the capital in one particular asset. We can also introduce different lending and borrowing rates by defining a “lending asset” (with $w_i \geq 0$) and a “borrowing asset” (with $w_j \leq 0$).

3.1.3 The Shape of the MV Frontier of Risky Assets: The Number of 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 3.5 for an example.

3.1.4 The Shape of the MV Frontier of Risky Assets: Correlations

This section discusses how the shape of the MV frontier depends on the correlation of the assets.

With intermediate correlations ($-1 < \rho < 1$) the mean-variance frontier is a hyperbola—see Figure 3.6. Notice that the mean–volatility trade-off improves as the correlation de-

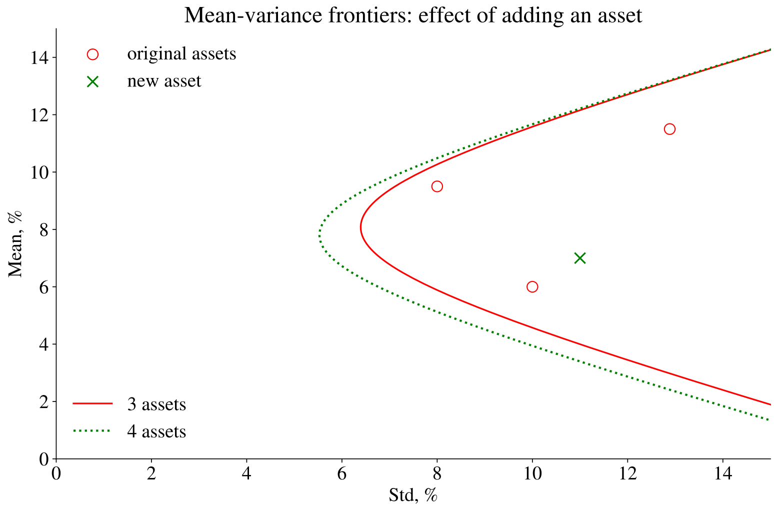


Figure 3.5: Mean-variance frontiers

creases: 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 Figure 3.7. 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 assets. The intuition is that a perfect correlation means that the second asset is a linear 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 again a pair of straight lines, see Figure 3.7. 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 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

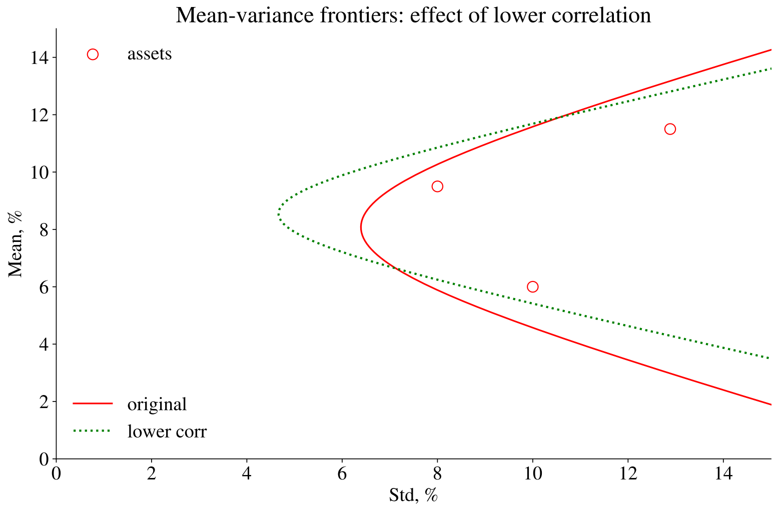


Figure 3.6: Mean-variance frontiers

sales). Then we get

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

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

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

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 $w_1 = \sigma_2/(\sigma_2 - \sigma_1)$.

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

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

There are two cases here (corresponding to the two lines mentioned in the text). First,

$$\sigma_p = \left\{[w_1\sigma_1 - (1-w_1)\sigma_2]^2\right\}^{1/2} = w_1\sigma_1 - (1-w_1)\sigma_2,$$

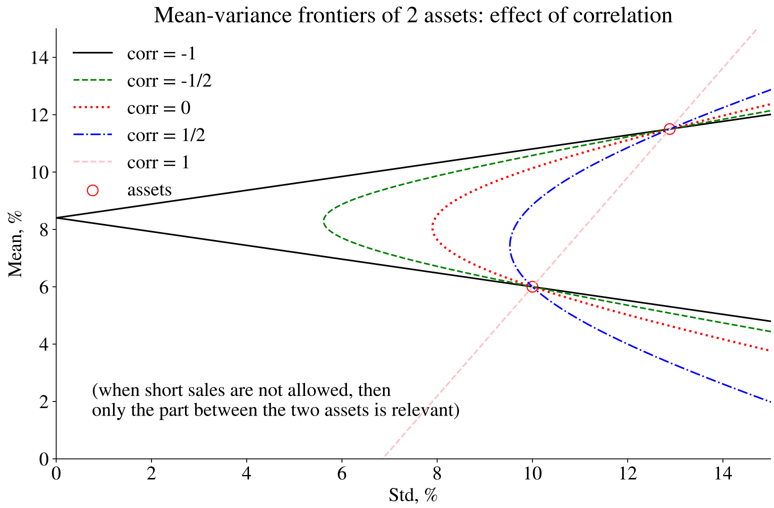


Figure 3.7: Mean-variance frontiers for two risky assets: different correlations. The two assets are indicated by circles. Points beyond the two assets can be generated by negative portfolio weights.

if the term in square brackets is positive. Second,

$$\sigma_p = \left\{[-w_1\sigma_1 + (1-w_1)\,\sigma_2]^2\right\}^{1/2} = -w_1\sigma_1 + (1-w_1)\,\sigma_2,$$

if the term in these square brackets is positive. Actually, the 2nd expression in brackets 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 $w_1 = \sigma_2/(\sigma_1 + \sigma_2)$. ■

3.1.5 Calculating the MV Frontier of Risky Assets

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 (warning: recycled notation) as

$$A = \mu'\Sigma^{-1}\mu,\ B = \mu'\Sigma^{-1}\mathbf{1},\ \text{and}\ C = \mathbf{1}'\Sigma^{-1}\mathbf{1}, \tag{3.7}$$

where $\mathbf{1}$ is a (column) vector of ones (as many as there are assets) and μ' is the transpose

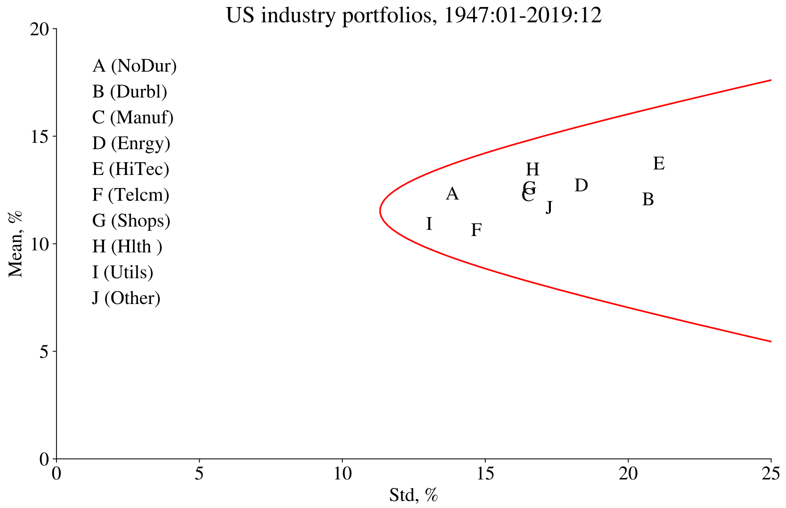


Figure 3.8: MV frontier from US industry indices

of the column vector μ. Then, calculate the scalars (for a given required return μ^*)

$$\lambda = \frac{C\mu^* - B}{AC - B^2} \text{ and } \delta = \frac{A - B\mu^*}{AC - B^2}. \tag{3.8}$$

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

$$w = \Sigma^{-1}(\mu\lambda + \mathbf{1}\delta). \tag{3.9}$$

Using this in $w'\Sigma w$ 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$\times$mean space, the efficient frontier (the upper part) is concave.

Remark 3.6 *(Scaling of returns I) Returns are often expressed in percents, that is, multiplied by 100. Clearly, this scales up means and standard deviations by 100, so e.g. the Sharpe ratio is the same as for the original returns. This carries over to many calculations. However, multiplying by 100 scales up the variance (and any covariance) by* 100^2*, so care is needed when using expressions that involve both means and variances (or covariances).*

Remark 3.7 *(Scaling of the returns II) The portfolio weights (3.9) are unchanged if both* μ *and* Σ *are scaled by (for instance) 52. Notice, however, that the average returns of the new MV portfolios are scaled by 52, but the standard deviations by just* $\sqrt{52}$.

Another way to construct the MV frontier of risky assets is to retrace it by *combining any two portfolios on the frontier*. For instance, we can use

$$\begin{aligned} w_\kappa &= \kappa w_g + (1-\kappa)\, w_T, \text{ where} \\ w_g &= \Sigma^{-1}\mathbf{1}/\mathbf{1}'\Sigma^{-1}\mathbf{1} \text{ and} \\ w_T &= \Sigma^{-1}\mu^e/\mathbf{1}'\Sigma^{-1}\mu^e. \end{aligned} \tag{3.10}$$

The first line defines a portfolio in terms of two portfolios (w_g and w_T) that are known to be on the MV frontier. The first (w_g) is the global minimum variance portfolio (lowest possible variance) and the second (w_T) is the tangency portfolio (to be discussed later on). The mean return can be calculated as $w_\kappa'\mu$ and the variance as $w_\kappa'\Sigma w_\kappa$. See Figure 3.11.

Proof. (of (3.7)–(3.9)) 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

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

(This can be written $\Sigma w - \lambda\mu - \delta\mathbf{1} = \mathbf{0}$.) 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)

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

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

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

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. ■

Remark 3.8 *(*Alternative way to calculate w) Combine the first order conditions in the proof of (3.7)–(3.9) as*

$$\begin{bmatrix} \Sigma & \mu & \mathbf{1} \\ \mu' & 0 & 0 \\ \mathbf{1}' & 0 & 0 \end{bmatrix} \begin{bmatrix} w \\ \tilde{\lambda} \\ \tilde{\delta} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mu^* \\ 1 \end{bmatrix},$$

where $\mathbf{0}$ is a vector of n zeros. Solve for the vector $(w, \tilde{\lambda}, \tilde{\delta})$. It can be shown that this gives the same results for w as (3.7)–(3.9).

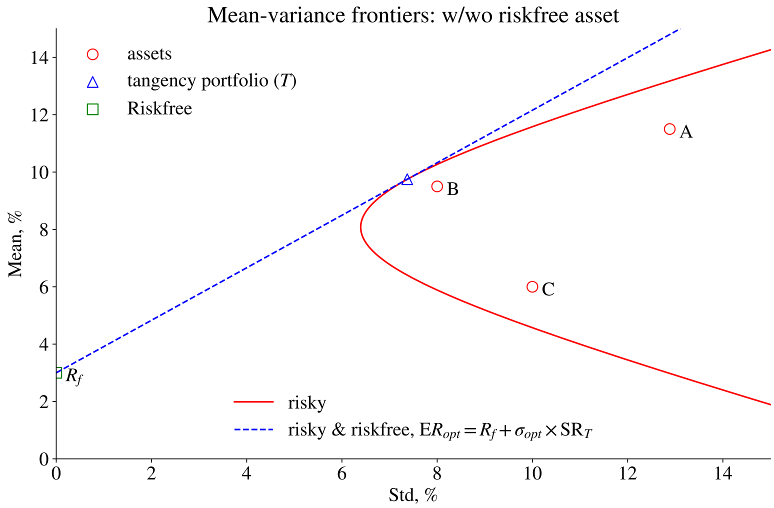


Figure 3.9: Mean-variance frontiers, w/wo riskfree asset

3.2 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

$$\begin{aligned} R_p &= w_1 R_1 + w_2 R_2 + (1 - w_1 - w_2) R_f \\ &= w_1 (R_1 - R_f) + w_2 (R_2 - R_f) + R_f \\ &= w_1 R_1^e + w_2 R_2^e + R_f, \end{aligned} \tag{3.11}$$

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

$$\begin{gathered} \min_{w_1, w_2} (w_1^2 \sigma_{11} + w_2^2 \sigma_{22} + 2 w_1 w_2 \sigma_{12})/2 \\ \text{subject to } w_1 \mu_1^e + w_2 \mu_2^e + R_f = \mu^*. \end{gathered} \tag{3.12}$$

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

$$\begin{gathered} \min_w w' \Sigma w \text{ subject to} \\ w' \mu^e + R_f = \mu^*. \end{gathered} \tag{3.13}$$

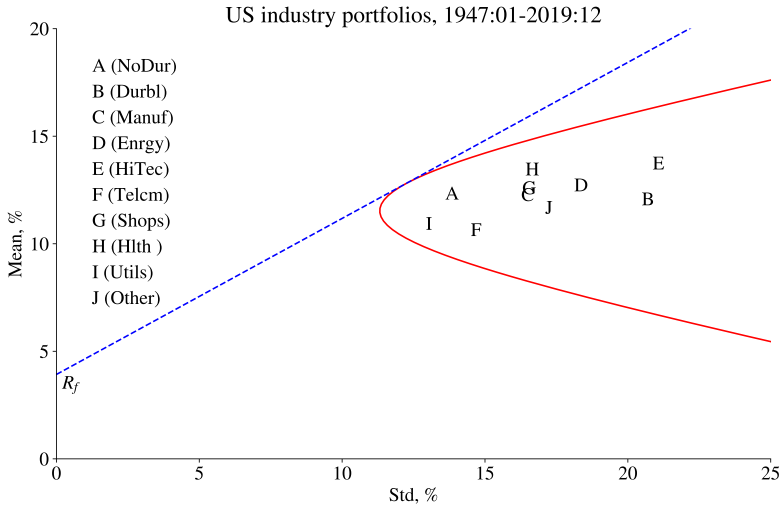


Figure 3.10: M-V frontier from US industry indices

When there are no additional constraints, then we can find an explicit solution in terms of some matrices and vectors—see Section 3.2.1. In all other cases, we need to apply an explicit numerical minimization algorithm (preferably for quadratic models).

3.2.1 Calculating the MV Frontier of Riskfree and Risky Assets

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

$$w = \frac{\mu^* - R_f}{(\mu^e)' \Sigma^{-1} \mu^e} \Sigma^{-1} \mu^e, \tag{3.14}$$

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 $w'\Sigma w$ 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 3.9 for an illustration and Figure 3.10 for an empirical example.

Proof. (of (3.14)) 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

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

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

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

(This can be written $\Sigma w - \lambda\mu^e = \mathbf{0}$.) We can solve for the portfolio weights as

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

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

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

Combine to get

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

Use in the above expression for w. ■

Remark 3.9 *(*Alternative way to calculate w) Combine the first order conditions in the proof of (3.14) as*

$$\begin{bmatrix} \Sigma & \mu^e \\ \mu^{e\prime} & 0 \end{bmatrix} \begin{bmatrix} w \\ \tilde{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mu^* - R_f \end{bmatrix}.$$

Solve for the vector $(w, \tilde{\lambda})$.

Remark 3.10 *(* $\min \operatorname{Std}(R_p)$ st $\operatorname{E} R_p = \mu^*$ and $\Sigma_{i=1}^n w_i = 1$) Consider the Lagrangian $(w'\Sigma w)^{1/2} + \tilde{\lambda}(\mu^* - w'\mu^e - R_f)$. The first order conditions are $(w'\Sigma w)^{-1/2}\Sigma w = \tilde{\lambda}\mu^e$ and $\mu^* = w'\mu^e + R_f$. Try the MV solution (3.14). We know that is satisfies $\mu^* = w'\mu^e + R_f$. We can also easily calculate $w'\Sigma w = (\mu^* - R_f)^2/(\mu^e)'\Sigma^{-1}\mu^e$. Using this and (3.14) in the first order condition shows that it holds with $\tilde{\lambda} = 1/\sqrt{(\mu^e)'\Sigma^{-1}\mu^e}$.*

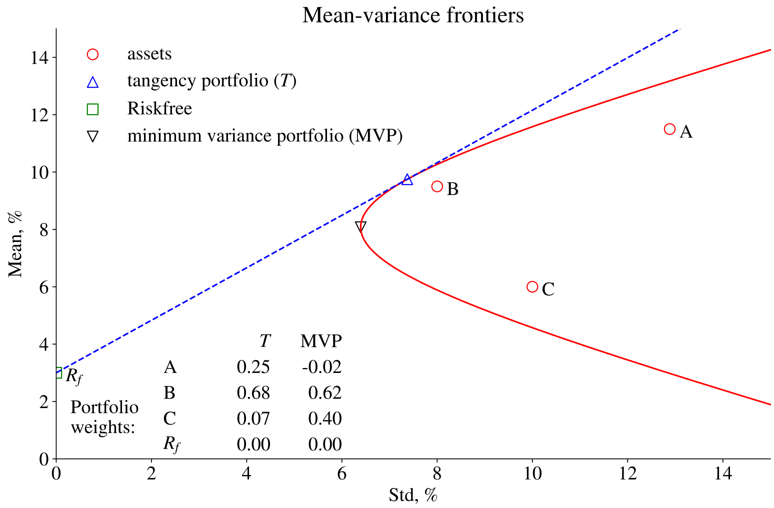


Figure 3.11: Mean-variance frontiers, tangency and MVP

3.3 The Tangency Portfolio

The MV frontier for risky assets and the frontier for risky + riskfree assets are tangent at one point—called the *tangency portfolio*: see Figures 3.9 and 3.11. In this case the portfolio weights (3.9) and (3.14) coincide. Therefore, the portfolio weights of the risky assets (3.14) 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 (3.14) gives the portfolio weights of the tangency portfolio

$$w_T = \frac{\Sigma^{-1}\mu^e}{\mathbf{1}'\Sigma^{-1}\mu^e}. \tag{3.15}$$

Proof. (of (3.15)) Put the sum of the portfolio weights in (3.14) 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 at a particular value of μ^* where

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

Using in (3.14) gives (3.15). ■

Notice that the tangency portfolio has the *highest possible Sharpe ratio*, that is, the most favourable reward-risk ratio.

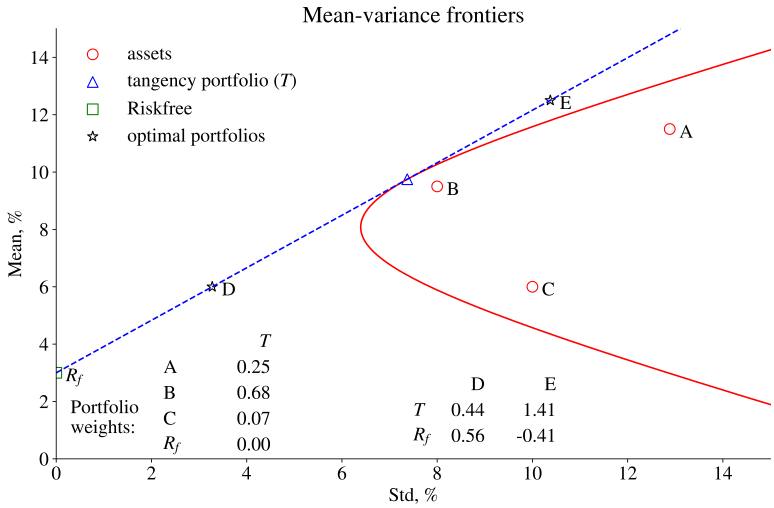


Figure 3.12: Mean-variance frontiers, creating portfolios by combining the tangency portfolio and the riskfree

Actually, every portfolio on the MV frontier (with risky assets and a riskfree asset) can be written

$$R_{opt} = vR_T + (1-v)R_f = vR_T^e + R_f, \tag{3.16}$$

where R_T is the return on the tangency portfolio. See Figure 3.12.

It follows that

$$\begin{aligned} \mathrm{E}\, R_{opt} &= v\, \mathrm{E}\, R_T^e + R_f \text{ and} \\ \mathrm{Var}(R_{opt}) &= v^2\, \mathrm{Var}(R_T). \end{aligned} \tag{3.17}$$

Combine the expressions for the mean and the variance to get (assuming $v \geq 0$) to get

$$\mathrm{E}\, R_{opt} = R_f + \mathrm{Std}(R_{opt}) SR_T, \tag{3.18}$$

where SR_T is the Sharpe ratio of the tangency portfolio. This is the equation for MV frontier including a riskfree asset (the straight line in the MV figures). It is also called the *capital market line.*

Remark 3.11 *(*Maximising the Sharpe ratio directly.) Maximizing $v'\mu^e/\sqrt{v'\Sigma v}$ gives the following n first order conditions*

$$\mu^e = \frac{v'\mu^e}{v'\Sigma v}\Sigma v.$$

It is straightforward to show that setting v equal to the tangency portfolio in (3.15) satisfy those first order conditions. (In particular, it helps to notice that $w_T' \mu^e / w_T' \Sigma w_T = \mathbf{1}' \Sigma^{-1} \mu^e$.) To be precise, any proportional scaling of the tangency portfolio ($v = \delta w_T$ where $\delta \neq 0$ is a scalar) will satisfy those first order conditions. This means any point on the capital market line. To find a unique solution, we therefore have to impose at least one restriction, for instance, that the portfolio weights v on the risky assets sum to 1.

Remark 3.12 *(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 $\operatorname{Var}(R_T^e) = \mu^{e\prime} \Sigma^{-1} \mu^e / \left(\mathbf{1}' \Sigma^{-1} \mu^e\right)^2$. It follows that $\mu_T^e / \operatorname{Var}(R_T^e) = \mathbf{1}' \Sigma^{-1} \mu^e$ and that the squared Sharpe ratio is $(\mu_T^e)^2 / \operatorname{Var}(R_T^e) = \mu^{e\prime} \Sigma^{-1} \mu^e$.*

3.3.1 Examples of Tangency Portfolios

Consider the simple case with two risky assets which are uncorrelated ($\sigma_{12} = 0$). The tangency portfolio (3.15) is then

$$\begin{bmatrix} w_{T,1} \\ w_{T,2} \end{bmatrix} = \frac{1}{\sigma_{22}\mu_1^e + \sigma_{11}\mu_2^e} \begin{bmatrix} \sigma_{22}\mu_1^e \\ \sigma_{11}\mu_2^e \end{bmatrix}. \tag{3.19}$$

This shows that if both excess returns are positive, then *(i)* the weight on asset i increases when μ_i^e increases and when σ_{ii} decreases; *(ii)* both weights are positive.

Example 3.13 *(Tangency portfolio, numerical) When $(\mu_1^e, \mu_2^e) = (8, 5)$, the correlation is zero, and $(\sigma_{11}, \sigma_{22}) = (256, 144)$, then (3.19) gives*

$$\begin{bmatrix} w_{T,1} \\ w_{T,2} \end{bmatrix} = \begin{bmatrix} 0.47 \\ 0.53 \end{bmatrix}.$$

When μ_1^e increases from 8 to 12, then we get

$$\begin{bmatrix} w_{T,1} \\ w_{T,2} \end{bmatrix} = \begin{bmatrix} 0.57 \\ 0.43 \end{bmatrix}.$$

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 (3.19) becomes

$$\begin{bmatrix} w_{T,1} \\ w_{T,2} \end{bmatrix} = \frac{1}{(\mu_1^e + \mu_2^e)(1-\rho)} \begin{bmatrix} \mu_1^e - \rho\mu_2^e \\ \mu_2^e - \rho\mu_1^e \end{bmatrix}. \tag{3.20}$$

Results: *(i)* both weights are positive if the returns are negatively correlated ($\rho < 0$) and both excess returns are positive; *(ii)* $w_{T,2} < 0$ 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 3.14 *(Tangency portfolio, numerical) When* $(\mu_1^e, \mu_2^e) = (8, 5)$*, and* $\rho = -0.8$ *we get*

$$\begin{bmatrix} w_{T,1} \\ w_{T,2} \end{bmatrix} = \begin{bmatrix} 0.51 \\ 0.49 \end{bmatrix}.$$

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

$$\begin{bmatrix} w_{T,1} \\ w_{T,2} \end{bmatrix} = \begin{bmatrix} 1.54 \\ -0.54 \end{bmatrix}.$$

Chapter 4

Index Models

Reference: Elton, Gruber, Brown, and Goetzmann (2014) 7–8

4.1 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

$$\begin{aligned} \mathrm{E}\, R_p &= \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \\ \mathrm{Var}(R_p) &= \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}. \end{aligned} \tag{4.1}$$

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

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}, \tag{4.2}$$

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

There are two main problems in estimating these parameters: (1) the number of parameters increase very quickly as the number of assets increases and (2) 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, it is often found that asset pairs with extreme (very low or very high) historical correlations, tend to have more normal correlations in future time periods.

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:

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

4.2 The Market Model

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_{mt})

$$R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it}, \text{ where} \tag{4.3}$$
$$\mathrm{E}\,\varepsilon_{it} = 0, \ \mathrm{Cov}(\varepsilon_{it}, R_{mt}) = 0.$$

In this regression R_{it} is the return on asset i in period t, while R_{mt} is the market return in the same period. The regression is done on time series (R_{it} and R_{mt} for $t = 1, 2, ..., T$). As usual, the regression slope is $\beta_i = \text{Cov}(R_i, R_m)/\text{Var}(R_m)$. This regression may use the net returns (as indicated above), or the returns in excess of a riskfree rate. The results for the β (which is the focus here) are typically very similar.

The 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 (4.3) is estimated by Least Squares.

Remark 4.1 *(Beta of a portfolio) For a portfolio of assets i and j, the beta is*

$$\beta_p = w_i \beta_i + w_j \beta_j.$$

(This follows from the fact that $\text{Cov}(w_i R_i + w_j R_j, R_m) = w_i \text{Cov}(R_i, R_m) + w_j \text{Cov}(R_j, R_m)$.*) In short, the beta of a portfolio is the portfolio of betas. This applies also to long-short portfolios where $w_i = 1$ and $w_j = -1$ (and where the weight do not sum to unity), and it gives $\beta_p = \beta_i - \beta_j$. Alternatively, to create a portfolio with a zero beta, you could buy $w_i = 1/\beta_i$ (in value terms) of asset i and $w_j = -1/\beta_j$ of asset j (short selling if $\beta_j > 0$). Such a portfolio should be uncorrelated with the market moves.*

Remark 4.2 *(Market indices I) A market index I_t is calculated as*

$$I_t = (1 + R_{mt}) I_{t-1}, \text{ where } R_{mt} = \textstyle\sum_{i=1}^{n} w_{i,t-1} R_{i,t},$$

where i denotes the n different components/assets (for instance, stocks) of the index. This is a capital weighted return index if (a) R_{it} is the return on holding asset i between $t-1$ and t; and (b) $w_{i,t-1}$ is the market capitalization of asset i (for instance, number of shares times the price per share) relative to the total market capitalization of all n assets—measured at the end of period $t-1$. Most big indices are of this sort If, instead, $R_{i,t}$ only includes the capital gain of holding asset i, then the index is a price index. In other cases, the weights reflect the market capitalization of the floats (those shares that are actively traded). In yet other cases the weights are the same across the assets (an equally weighted index).

Remark 4.3 *(Market indices II*) Dow Jones Industrial Average and Nikkei 225 have very special weights. In practice, these two indices are just the average prices of all (30*

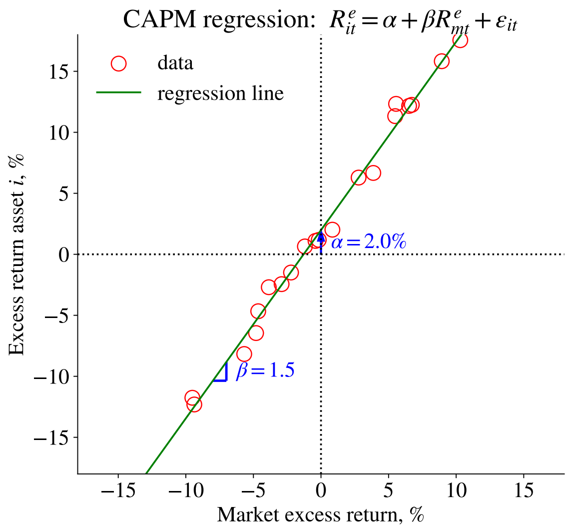


Figure 4.1: CAPM regression

or 225) stocks in the index. This means that the portfolio weights are proportional to the stock price.

Remark 4.4 *(Market indices III) More recently, a large number of alternative indices have been introduced, for instance of (a) "sustainable" companies (DJSI); (b) fundamentally weighted indices (weights based on sales, earnings or dividends); (c) 1/volatility based indices; (d) performance based indices (large weights on recent winners).*

See Figures 4.1–4.3 for illustrations. See also Figure 4.4 for some alternative assets.

4.3 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_{mt}). This means that we add one assumption to (4.3)

$$\text{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0, \tag{4.4}$$

which says that 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.

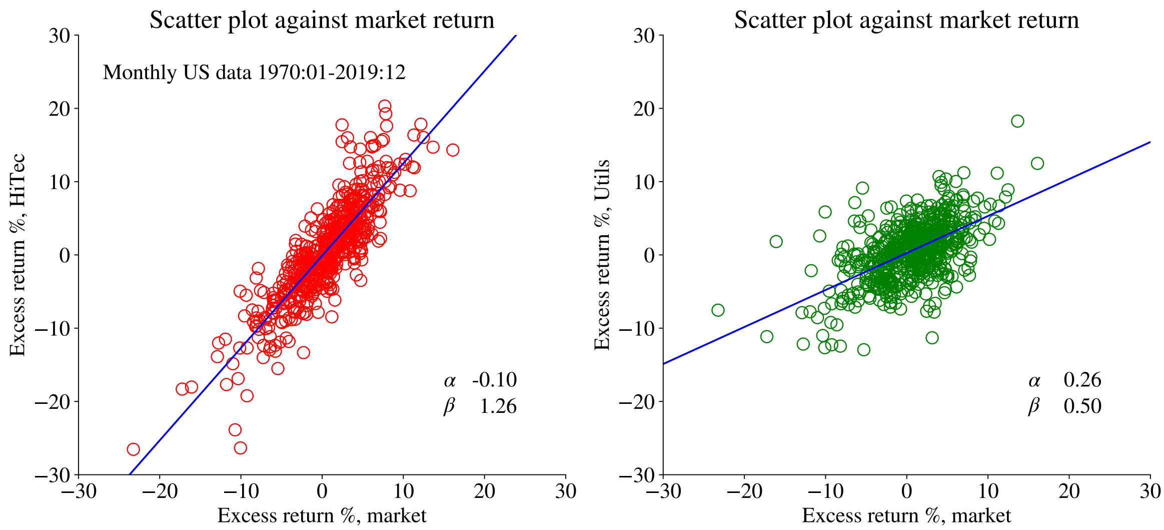


Figure 4.2: Scatter plot against market return

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.

If (4.3) and (4.4) are true, then the variance of asset i and the covariance of assets i and j are

$$\sigma_{ii} = \beta_i^2 \operatorname{Var}(R_{mt}) + \operatorname{Var}(\varepsilon_{it}) \tag{4.5}$$

$$\sigma_{ij} = \beta_i \beta_j \operatorname{Var}(R_{mt}) \text{ when } i \neq j. \tag{4.6}$$

Together, these equations show that we can calculate the whole covariance matrix by having just the variance of the index (to get Var (R_m)) and the output from n regressions (to get β_i and Var (ε_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).

See Figure 4.5 for an example based on the Fama-French portfolios detailed in Table 4.1.

Example 4.5 *(Two assets) Let* $[\beta_1, \beta_2] = [0.9, 1.1]$, $[\operatorname{Var}(\varepsilon_{1t}), \operatorname{Var}(\varepsilon_{2t})] = [100, 25]/10^2$,

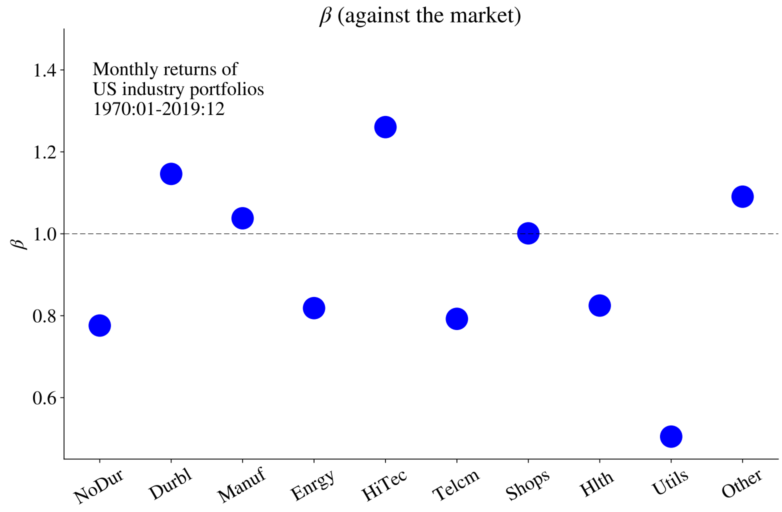


Figure 4.3: βs of US industry portfolios

and $\text{Var}\,(R_{mt}) = 225/100^2$. *Then*

$$\text{Cov}(R_t) \approx \begin{bmatrix} 282.25 & 222.75 \\ 222.75 & 297.25 \end{bmatrix} /100^2.$$

Remark 4.6 *(Fama-French portfolios) The portfolios in Table 4.1 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 (also called double sorting). For instance, in Table 4.1 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.*

Proof. (of (4.5)–(4.6) By using (4.3) and (4.4) and recalling that $\text{Cov}(R_m, \varepsilon_i) = 0$ direct calculations give

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

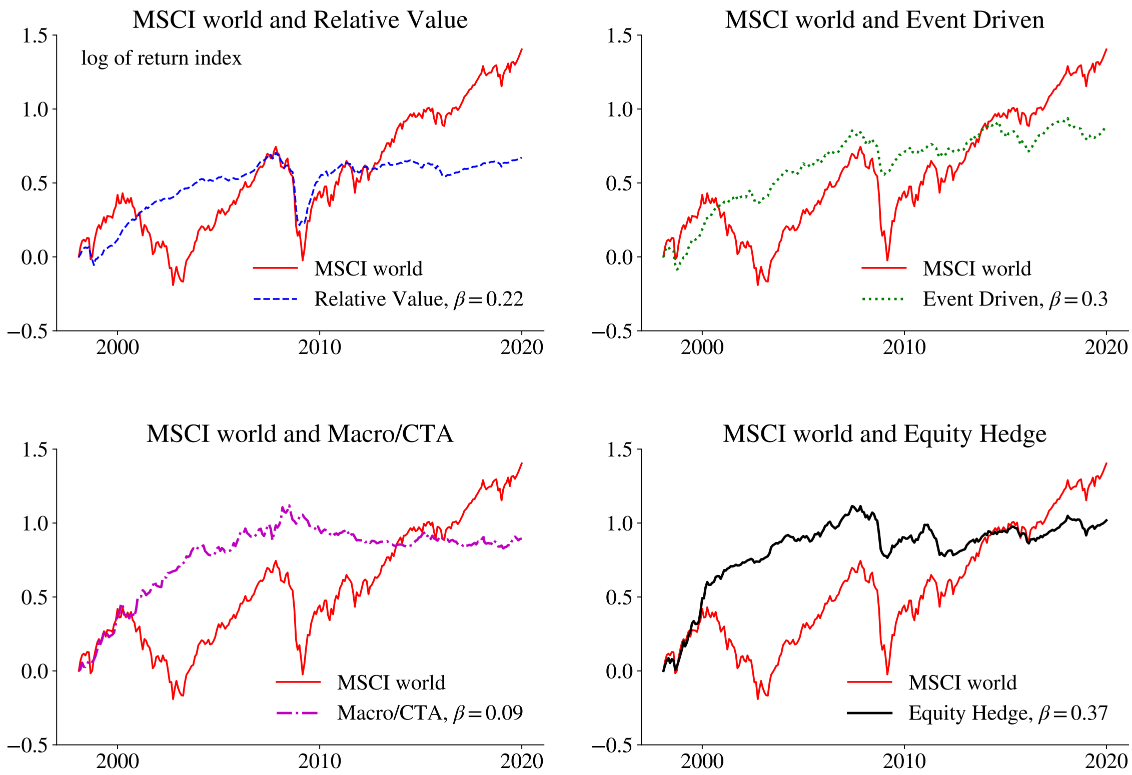


Figure 4.4: Comparing hedge fund indices with the MSCI world (equity) index

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

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

■

4.4 Estimating Beta

4.4.1 Estimating Historical Beta: OLS and Other Approaches

Least Squares (LS) is typically used to estimate α_i, β_i and Std(ε_{it}) in (4.3)—and the R^2 is used to assess the quality of the regression.

To assess the accuracy of historical betas, Blume (1971) and others estimate betas for

	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 4.1: Numbering of the FF portfolios.

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

$$\hat{\beta}_{i2} = \gamma_0 + \gamma_1 \hat{\beta}_{i1} + \upsilon_i \tag{4.7}$$

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 (4.3). 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

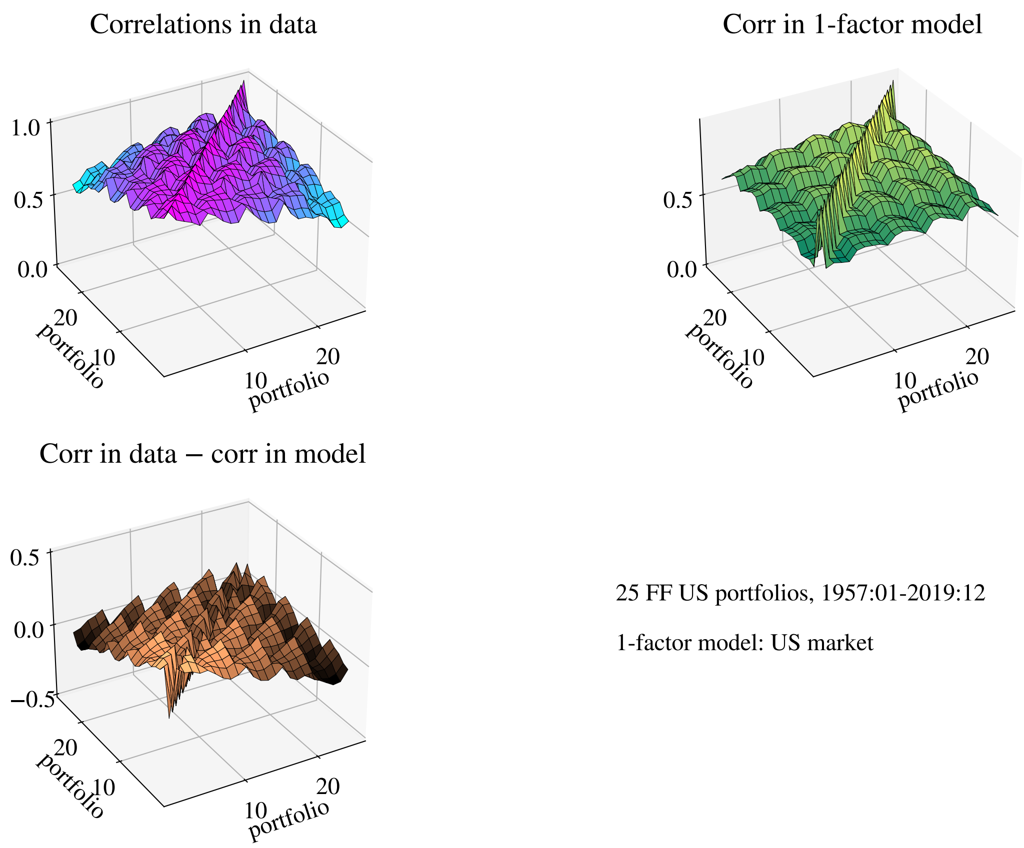


Figure 4.5: Correlations of US portfolios, single-index model

estimator of beta is

$$b = (1 - F)\hat{\beta}_{i1} + F\beta_0, \text{ where}$$
$$F = \frac{1/\sigma_0^2}{1/\sigma_0^2 + 1/\sigma_{\beta 1}^2} = \frac{\sigma_{\beta 1}^2}{\sigma_0^2 + \sigma_{\beta 1}^2}. \tag{4.8}$$

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.

4.4.2 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.

4.5 Multi-Index Models

4.5.1 Overview

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

$$\begin{aligned} R_{it} &= a_i + b_i' I_t + \varepsilon_{it}, \text{ where} \qquad (4.9) \\ \mathrm{E}\, \varepsilon_{it} &= 0,\ \mathrm{Cov}(\varepsilon_{it}, I_t) = \mathbf{0}, \text{ and } \mathrm{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0. \end{aligned}$$

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. Industry indices are also commonly used.

It turns out (see below) that the calculations of the covariance matrix are simpler if the indices are transformed to be uncorrelated.

Remark 4.7 *(Fama-French factors) Fama and French (1993) use three factors: the market excess 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 a high ratio of book value to market value minus the return on a portfolio with a low ratio (HML). All three are excess returns (although only the first is in excess of a riskfree return), since they are long-short portfolios.*

4.5.2 "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, $I_{1t}^* = I_{1t}$ (possibly demeaned). This would often be the market return.

2. Regress the second original index on the first transformed index, $I_{2t} = \gamma_0 + \gamma_1 I_{1t}^* + \varepsilon_{2t}$. Then, let the second transformed index be the intercept plus the fitted residual, $I_{2t}^* = \gamma_0 + \hat{\varepsilon}_{2t}$.

3. Regress the third original index on the first two transformed indices, $I_{3t} = \theta_0 + \theta_1 I_{1t}^* + \theta_2 I_{2t}^* + \varepsilon_{3t}$. Then, let $I_{3t}^* = \theta_0 + \hat{\varepsilon}_{3t}$. Follow the same idea for all subsequent indices.

Recall that the fitted residual (from Least Squares) is uncorrelated with the regressor (by construction). In this case, this means that I_{2t}^* is not correlated with I_{1t}^* (step 2) and that I_{3t}^* is not correlated with either I_{1t}^* and I_{2t}^* (step 3). The correlation matrix of the first three rotated indices is therefore

$$\text{Corr}\left(\begin{bmatrix} I_{1t}^* \\ I_{2t}^* \\ I_{3t}^* \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.10}$$

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. Notice that no information is lost in these transformations: the R^2 of the index model with rotated indices is the same as the R^2 from the model using the original indices.

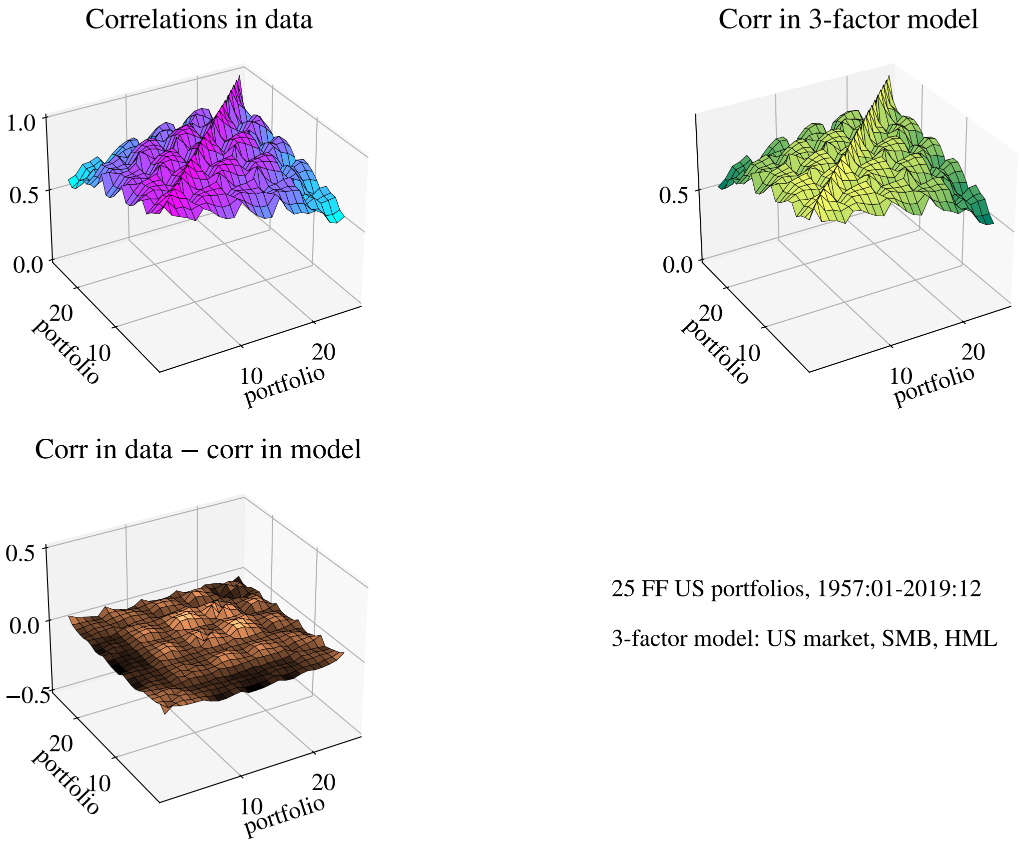


Figure 4.6: Correlations of US portfolios, multi-index model

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

$$I^*_{kt} = \delta_k + \hat{\varepsilon}_{kt}, \tag{4.11}$$

where δ_k and $\hat{\varepsilon}_k$ are the fitted intercept and residual from the regression

$$I_{kt} = \delta_k + \textstyle\sum_{s=1}^{k-1} \gamma_{ks} I^*_{st} + \varepsilon_{kt}. \tag{4.12}$$

Notice that for the first index ($k = 1$), the regression is only $I_1 = \delta_1 + \varepsilon_1$, so I^*_1 equals I_1.

4.5.3 Using the Multi-Index Model

If Ω is the covariance matrix of the indices, then the covariance of assets i and j is

$$\sigma_{ii} = b_i' \Omega b_i + \text{Var}(\varepsilon_{it}), \tag{4.13}$$

$$\sigma_{ij} = b_i' \Omega b_j \text{ when } i \neq j, \tag{4.14}$$

where b_i is the vector of slope coefficients obtained from regressing R_{it} on the vector of factors (I_t or I_t^*) as in (4.9). See Figure 4.5 for an example.

Remark 4.8 *(With uncorrelated factors) In case the factors are uncorrelated, then Ω is diagonal so the $b_i' \Omega b_j$ in (4.14) can be simplified to $\sum_{k=1}^{K} b_{ik} b_{jk} \text{Var}(I_{kt})$, where b_{ik} is the coefficient on factor k in the regression of R_{it}.*

4.5.4 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.

In fact, 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). The implications of those opinions for individual assets depend on the factor loadings.

4.6 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 judgemental adjustment of the historical excess return. Evidence suggest that 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) 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 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.)

Chapter 5

Portfolio Choice

Reference: Danthine and Donaldson (2005) 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 only risky assets 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 how the utility is maximized by moving as far to the upper left as possible—while staying in the set of feasible portfolios (on or below the mean-variance frontier). Also, see Figure 5.2 for an illustration of how the attitude towards risk determines which point of the mean-variance frontier that is optimal. The purpose of this section is to study the tangency portfolio is some detail.

5.1.1 A Risky Asset and a Riskfree Asset (recap)

Suppose there are one risky asset (i) 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*

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

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

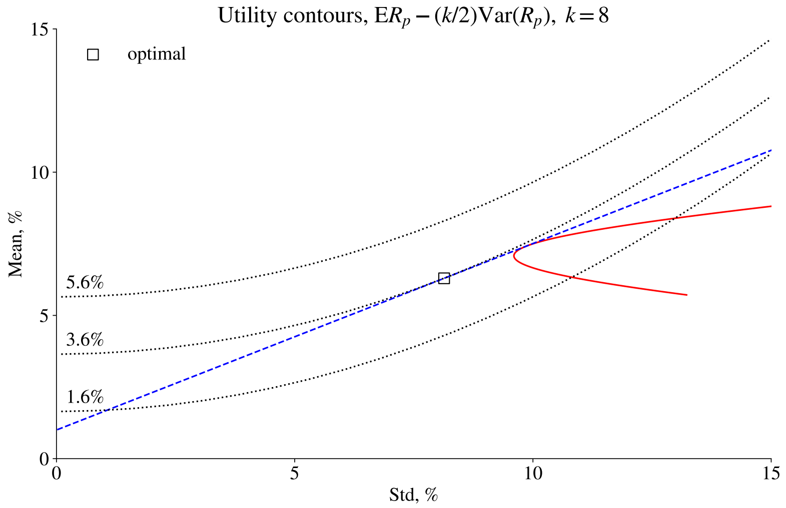


Figure 5.1: Iso-utility curves, mean-variance utility

(Dividing k by 2 is just a normalization of risk aversion: it makes the equation for the optimal portfolio choice look a bit less involved.) We have already demonstrated that the optimal portfolio weight of the risky asset is

$$v = \frac{1}{k} \frac{\mu_i^e}{\sigma_{ii}}. \tag{5.3}$$

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. The portfolio weight on the riskfree asset is $1 - v$.

Example 5.1 *(Portfolio choice) If $\mu_i^e = 3$, $\sigma_{ii} = 9$ and $k = 0.5$, then $v \approx 0.67$. Instead, with $k = 0.25$, $v \approx 1.33$.*

We have also show that the optimal solution implies that

$$\frac{\mathrm{E}\, R_{opt}^e}{\text{Var}(R_{opt})} = k, \tag{5.4}$$

where R_{opt} 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.

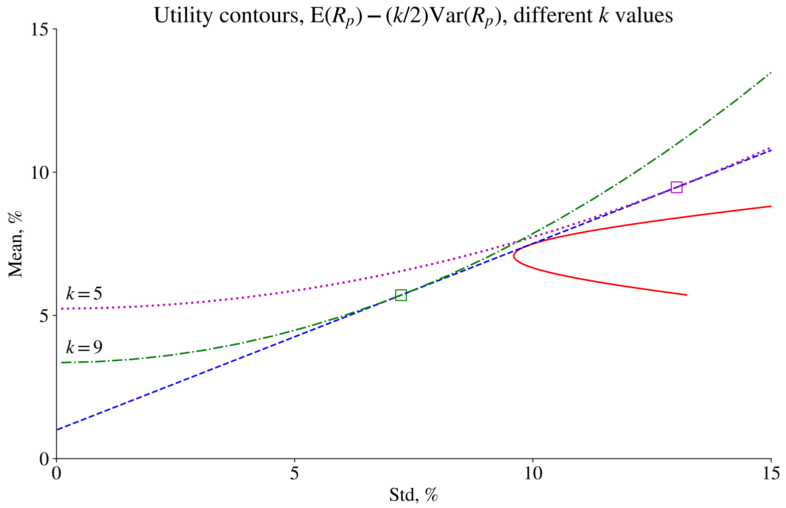


Figure 5.2: Iso-utility curves, mean-variance utility (different risk aversions)

5.2 Portfolio Choice with Several Risky Assets and a Riskfree Asset

5.2.1 The Optimization Problem

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_1, v_2) to maximize

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

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

Example 5.2 *(Optimal portfolio weights) Figure 5.3 shows the properties of two risky assets and a riskfree—and solves for the optimal portfolio weights (and also the tangency portfolio).*

Combining gives

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

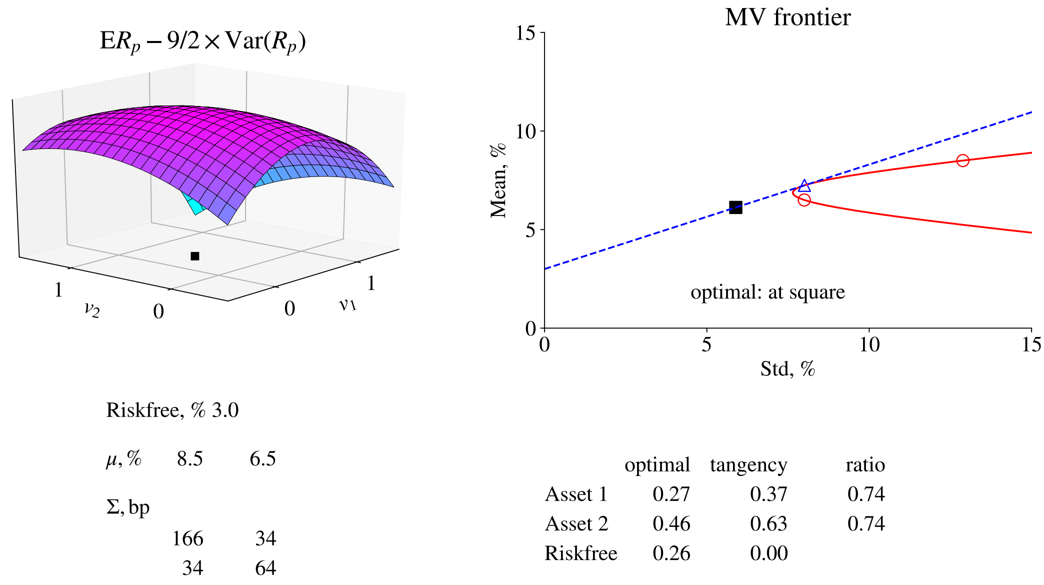


Figure 5.3: Choice of portfolios weights

where σ_{12} denotes the covariance of asset 1 and 2. In terms of the vector of portfolio weights (v), the same equation is

$$\mathrm{E}\,U(R_p) = v'\mu^e + R_f - \frac{k}{2}v'\Sigma v, \tag{5.8}$$

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

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

$$\begin{bmatrix} \partial\,\mathrm{E}\,U(R_p)/\partial v_1 \\ \partial\,\mathrm{E}\,U(R_p)/\partial v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{5.9}$$

which can be solved as

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \frac{1}{k}\frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}\begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}\begin{bmatrix} \mu_1^e \\ \mu_2^e \end{bmatrix} \text{ or} \tag{5.10}$$

$$v = \frac{1}{k}\Sigma^{-1}\mu^e. \tag{5.11}$$

Notice that the weight on the riskfree asset is $1 - \mathbf{1}'v$, where $\mathbf{1}$ is a (column) vector of ones.

Proof. (of (5.11)) The first order conditions are

$$
\begin{aligned}
0 &= \partial \operatorname{E} U(R_p)/\partial v_1 = \mu_1^e - \frac{k}{2}(2v_1\sigma_{11} + 2v_2\sigma_{12}) \\
0 &= \partial \operatorname{E} U(R_p)/\partial v_2 = \mu_2^e - \frac{k}{2}(2v_2\sigma_{22} + 2v_1\sigma_{12})\text{, or} \\
\begin{bmatrix} 0 \\ 0 \end{bmatrix} &= \begin{bmatrix} \mu_1^e \\ \mu_2^e \end{bmatrix} - k \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \\
\mathbf{0}_{2\times1} &= \mu^e - k\Sigma v.
\end{aligned}
$$

We can solve this linear system of equations as (5.11). ■

Remark 5.3 *(Portfolio choice with two risky assets) When there are only two risky assets, then (5.10) can be used to notice that*

$$v_1 > 0 \text{ if } \mu_1^e/\sigma_1 > \rho\mu_2^e/\sigma_2.$$

(Switch the subscripts to get a similar expression for v_2.) This shows that 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. See Figure 5.4 for an empirical illustration. (To derive this result, notice that the denominator $(\sigma_{11}\sigma_{22} - \sigma_{12}^2)$ in (5.10) is positive—since correlations are between -1 and 1. Then use the fact that $\sigma_{12} = \rho\sigma_1\sigma_2$ where ρ is the correlation coefficient.)

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

$$\frac{\operatorname{E} R_{opt}^e}{\operatorname{Var}(R_{opt})} = k. \tag{5.12}$$

This just says that higher risk aversion tilts the portfolio away from a high variance.

Proof. (of (5.12)) Use the portfolio weights in (5.11) to write

$$
\begin{aligned}
\frac{\operatorname{E} R_{opt}^e}{\operatorname{Var}(R_{opt})} = \frac{v'\mu^e}{v'\Sigma v} &= \frac{\left(\frac{1}{k}\Sigma^{-1}\mu^e\right)'\mu^e}{\left(\frac{1}{k}\Sigma^{-1}\mu^e\right)'\Sigma\left(\frac{1}{k}\Sigma^{-1}\mu^e\right)} \\
&= k\frac{\left(\Sigma^{-1}\mu^e\right)'\mu^e}{\left(\Sigma^{-1}\mu^e\right)'\mu^e} = k
\end{aligned}
$$

■

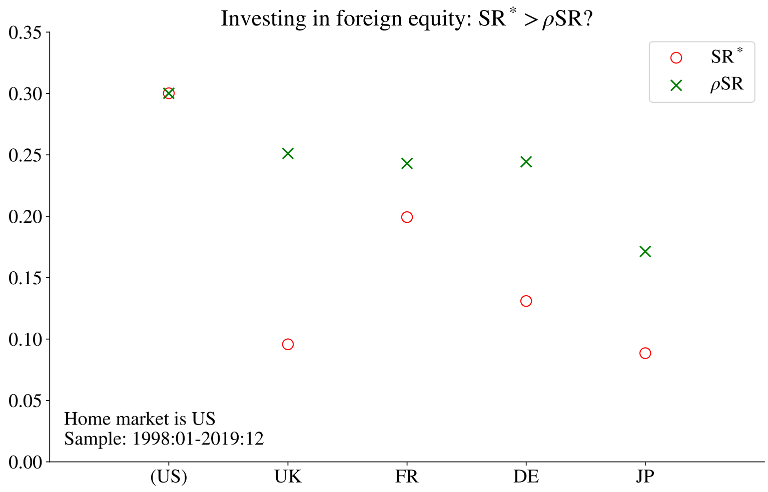


Figure 5.4: International stock indices

Remark 5.4 *(Several risky assets, no riskfree*) Maximizing the Lagrangian* $v'\mu - \frac{k}{2}v'\Sigma v + \theta(1 - v'\mathbf{1})$ *where* θ *is a Lagrange multiplier (for the constraint that the portfolio weights sum to one) gives the first order conditions (wrt.* v*)* $\mu - k\Sigma v - \theta\mathbf{1} = 0$*, which we can solve as* $v = \Sigma^{-1}(\mu + \theta\mathbf{1})/k$*. Combine this with the restriction to get* $1 = \mathbf{1}'\Sigma^{-1}(\mu + \theta\mathbf{1})/k$*, which gives* $\theta = (k - B)/C$*, where* $B = \mu'\Sigma^{-1}\mathbf{1}$*, and* $C = \mathbf{1}'\Sigma^{-1}\mathbf{1}$*. Finally, use this to substitute for* θ *to get* $v = \Sigma^{-1}[\mu/k + \mathbf{1}(1 - B/k)/C]$.

5.2.2 $\mathrm{E}\,R_p - \frac{k}{2}\,\mathrm{Var}(R_p)$ Gives a Portfolio on the Mean-Variance Frontier

It is clear that the optimal portfolio (5.11) is a scaling up/down of the *tangency portfolio* (see previous chapters)

$$w_T = \frac{\Sigma^{-1}\mu^e}{\mathbf{1}'\Sigma^{-1}\mu^e}. \tag{5.13}$$

Recall that the tangency portfolio has the highest Sharpe ratio, $\mathrm{E}\,R_p^e/\,\mathrm{Std}(R_p)$, of all portfolios.

Example 5.5 *(Tangency portfolio) Figure 5.3 shows also the tangency portfolio.*

We can use (5.13) to write the optimal portfolio weights in (5.11) as

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

where the first term is just a scalar. The balance $(1 - \mathbf{1}'v)$ is made up by a position in the riskfree asset. Note that all investors (different k, but same expectations) hold a mix of the tangency portfolio and the riskfree asset. They only differ in terms of how they mix. This *two-fund separation theorem* is very useful and has strong implications for portfolio choice. See Figure 5.5 for an illustration based on the asset characteristics in Table 11.1.

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 (5.14) shows that *(a)* when $k = \mathbf{1}'\Sigma^{-1}\mu^e$, 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 CML: *(c)* when $k > \mathbf{1}'\Sigma^{-1}\mu^e$ then the investor holds both the riskfree asset and the tangency portfolio in positive amounts and is on the CML to the left of the tangency portfolio, but *(d)* when $k < \mathbf{1}'\Sigma^{-1}\mu^e$ then the investor borrows at the riskfree rate and holds more than 100% of her wealth in the tangency portfolio (a leveraged) position.

	μ, %	Σ, bp		
		A	B	C
A	11.5	166	34	58
B	9.5	34	64	4
C	6.0	58	4	100

Table 5.1: Characteristics of the assets in the MV examples. Notice that μ, % is the expected return in % (that is, $\times 100$) and Σ, bp is the covariance matrix in basis points (that is, $\times 100^2$).

Another way to phrase the result in (5.14) is that we can *trace out the efficient set* (of risky and riskfree assets) by solving the portfolio choice problem for a range of different values of the the risk aversion (k). This is true also for the case without a riskfree asset (see the remark below).

Remark 5.6 *(Portfolio choice and the MVF, no riskfree*) The MV portfolio at the average return μ^* is $w = \Sigma^{-1}(\mu\lambda + \mathbf{1}\delta)$, where $\delta = (A - B\mu^*)/(AC - B^2)$ and where $A = \mu'\Sigma^{-1}\mu$, $B = \mu'\Sigma^{-1}\mathbf{1}$, and $C = \mathbf{1}'\Sigma^{-1}\mathbf{1}$ (the expression for λ is not important here). It is straightforward to show that setting $k = \lambda$ in Remark 5.4 gives the same portfolio weigths. This means that every portfolio in the efficient set can be generated by some k value.*

Remark 5.7 *(The mathematics of why $\max \mathrm{E}\, R_p - \frac{k}{2}\mathrm{Var}(R_p)$ gives a MV portfolio*) The efficient set solves the problem $\max \mathrm{E}\, R_p$ subject to $\mathrm{Var}(R_p) \leq c$ (where we vary c to trace out the efficient set). Notice that maximizing $\mathrm{E}\, R_p - \frac{k}{2}\mathrm{Var}(R_p)$ can be thought of as the Lagrangian formulation of this second problem.*

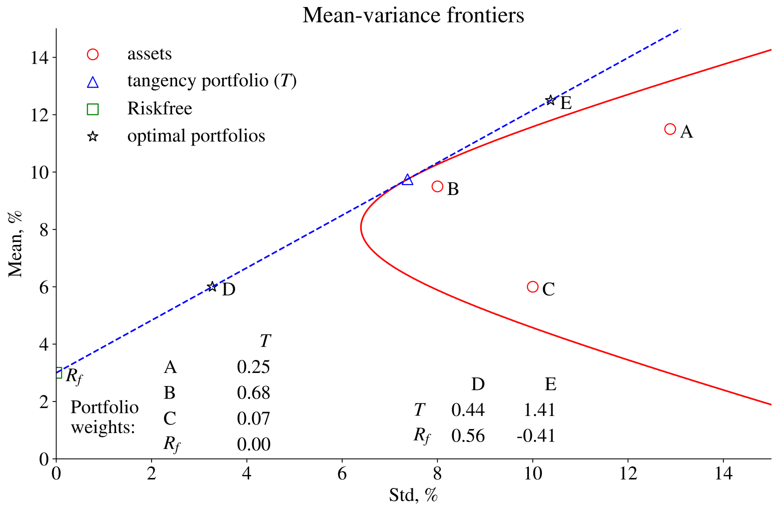


Figure 5.5: Mean-variance frontiers. The properties of the investable assets (A, B, and C) are shown in Table 11.1.

5.2.3 A Risky Asset and a Riskfree Asset Revisited

Once we have the tangency portfolio (with weights w_T as in (5.13)), 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 $R_T^e = w_T' R^e$ as R_i^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)

$$v^* = \frac{1}{k}\mathbf{1}'\Sigma^{-1}\mu^e, \tag{5.15}$$

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

Proof. (of (5.15)) Use the properties of the tangency portfolio (that $\mu_T^e/\sigma_T^2 = \mathbf{1}'\Sigma^{-1}\mu^e$) in equation (5.3),

$$v = \frac{1}{k}\frac{\mu_T^e}{\sigma_T^2} = \frac{1}{k}\mathbf{1}'\Sigma^{-1}\mu^e,$$

which is (5.15). It is a scalar since $\mathbf{1}'$ is $1 \times n$ and $\Sigma^{-1}\mu^e$ is $n \times 1$. ■

We can thus write all optimal portfolios (those on the CLM) as

$$R_p = vR_T + (1 - v)R_f. \tag{5.16}$$

This shows that we can easily construct *a portfolio with a desired beta*, since

$$\beta_p = v. \tag{5.17}$$

In practice, we can mix a (mutual or exchange traded) fund with cash to get whatever portfolio beta we want.

Example 5.8 *(Portfolio choice to get a desired β) To construct a portfolio with $\beta = 1.2$ against the tangency portfolio, invest $v = 1.2$ in the tangency portfolio and -0.2 in the riskfree.*

Proof. (of (5.17)) Recall that $\beta_p = \text{Cov}(R_p, R_T)/\text{Var}(R_T)$. Use (5.16) to write to the numerator as $\text{Cov}(vR_T, R_T) = v\,\text{Var}(R_T)$, since the riskfree has no covariance. Combine to get (5.17). ■

5.3 Historical Estimates of the Average Returns and the Covariance Matrix

Figure 5.6 illustrates mean returns and standard deviations, estimated by exponentially weighted moving averages (as by RiskMetrics). This means that the estimates are based on a longer and longer sample, but that old data are given lower weights. (In a sample that ends in t, the data in $t - s$ is given the weight λ^{t-s}, where $\lambda < 1$.)

Figure 5.8 shows how the optimal portfolio weights (based on mean-variance preferences) change over time. 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.

This means that, in practical application of the MV framework, we typically put restrictions on the levels and changes of the portfolio weights, see Figure 5.9 for an example.

5.4 Appendix: A Primer on Using Numerical Optimization Routines*

Reference: Brandimarte (2006), Stan manual (`http://mc-stan.org/users/documentation/`)

5.4.1 Unconstrained Minimization

Consider the loss function

$$f(\theta) = (x - 2)^2 + (4y + 3)^2, \tag{5.18}$$

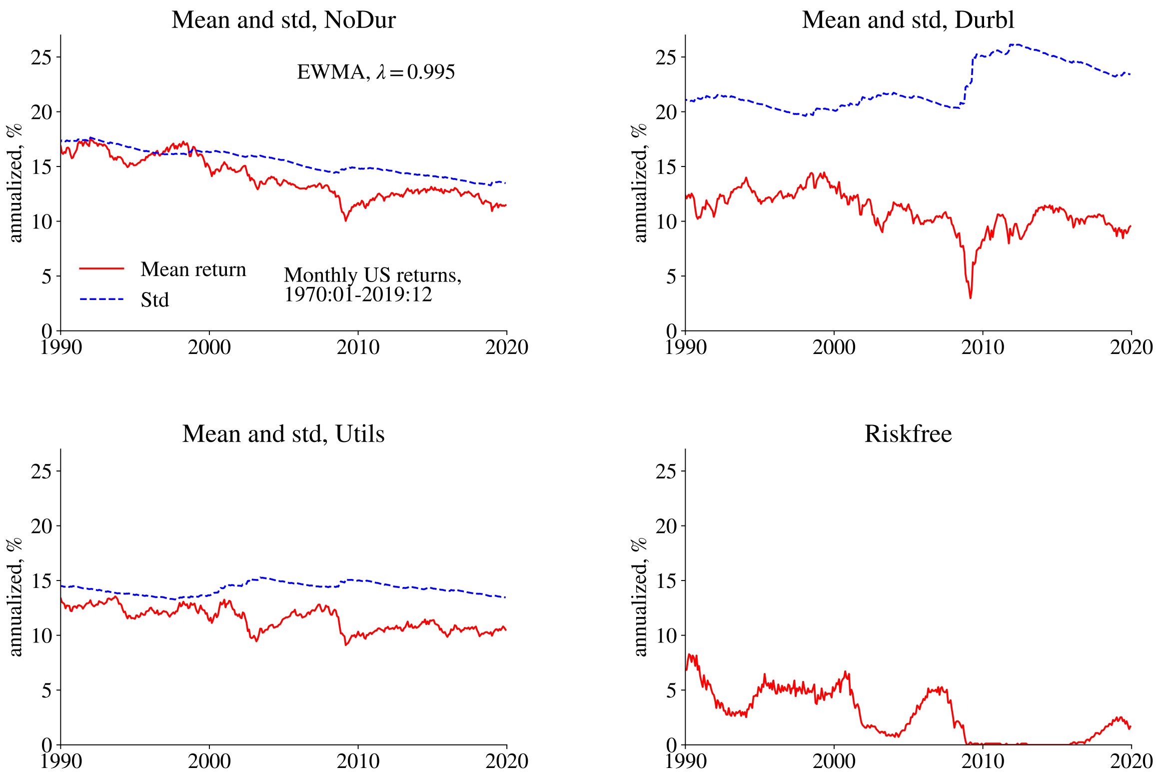


Figure 5.6: Dynamically updated estimates, 3 U.S. industries

where $\theta = (x, y)$ contains the two choice variables. (Clearly, the minimum of $f(\theta)$ is at $(x, y) = (2, -3/4)$.) Since this loss function is particularly simple (quadratic and also separable in x and y) the solutions below will be straightforward. However, the methods presented can also be used with more complicated loss functions.

A numerical minimization routine searches different values of θ, typically starting from a guess supplied by the user, to find the values that makes $f(\theta)$ as small as possible. Convergence criteria (often set by the user) determine when the search will stop (for instance, when the improvement in $f(\theta)$ is smaller than a certain threshold or when the θ values do not change much anymore). The starting guess is often important, so be sure to use reasonable values. See Figure 5.10 for an example.

There are two main types of algorithms: those that use derivatives of the loss function (which needs to be coded by the user) and those that do not ("derivative free"). The latter type is often slower, but sometimes more robust.

Most optimization algorithms are for minimizing a function value. In case you want to maximize, then just change the sign of the function and then minimize it. For instance,

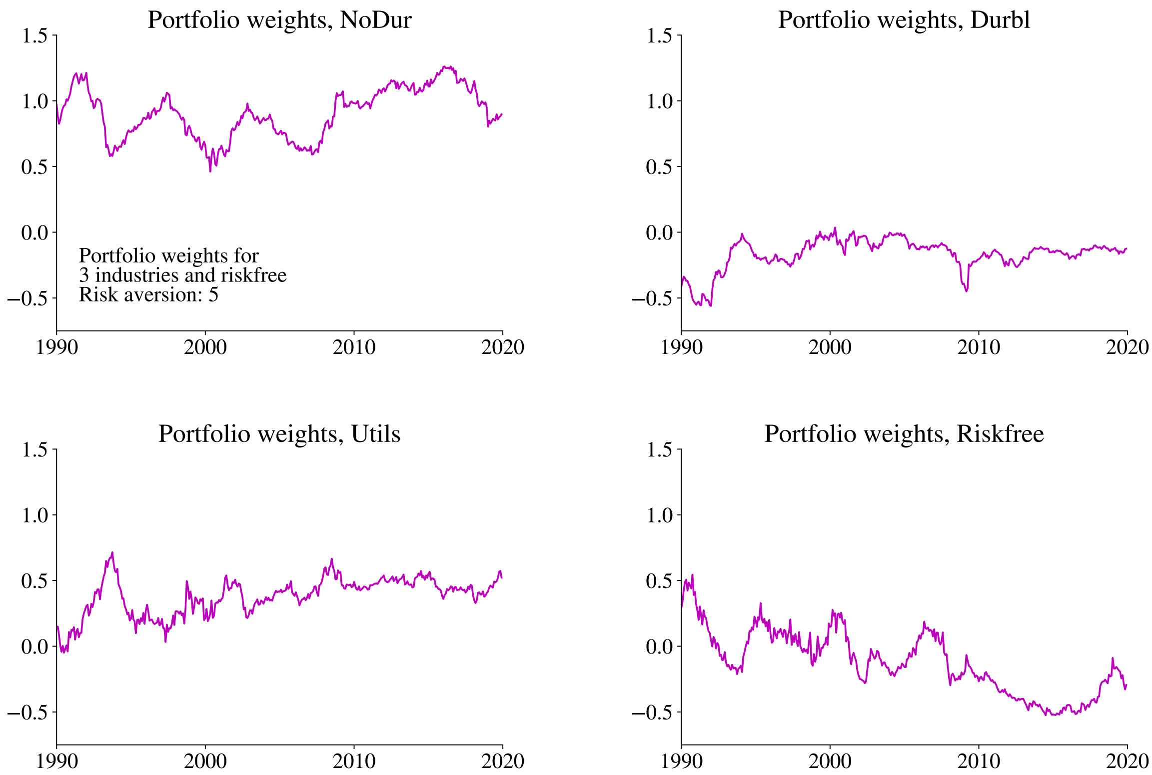


Figure 5.7: Dynamically updated portfolio weights, T-bill and 3 U.S. industries

if you want to maximize $g(\theta)$, then you can do that by minimizing $-g(\theta)$.

5.4.2 Bounds on Variables

Many numerical optimization packages have options for setting bounds on the solution ("box minimization"). As an alternative, we could transform the variables and then apply an algortihm for unconstrained optimisation. The latter is briefly discussed below.

A simple way to handle a lower bound like $a \leq x$ is to let the optimization routine optimize (without any restrictions) with respect to a transformed variable $\tilde{x} = \ln(x - a)$. Inside the loss function (and also after having obtained the minimizer) we transform back to x by $x = \exp(\tilde{x}) + a$.

Instead, with an upper bound $x \leq b$, we instead optimize over $\tilde{x} = \ln(b - x)$ and transform back to x by $x = b - \exp(\tilde{x})$.

Suppose we use the same loss function (5.18) as before, but also impose the bounds

$$2.75 \leq x \text{ and } y \leq -0.3. \tag{5.19}$$

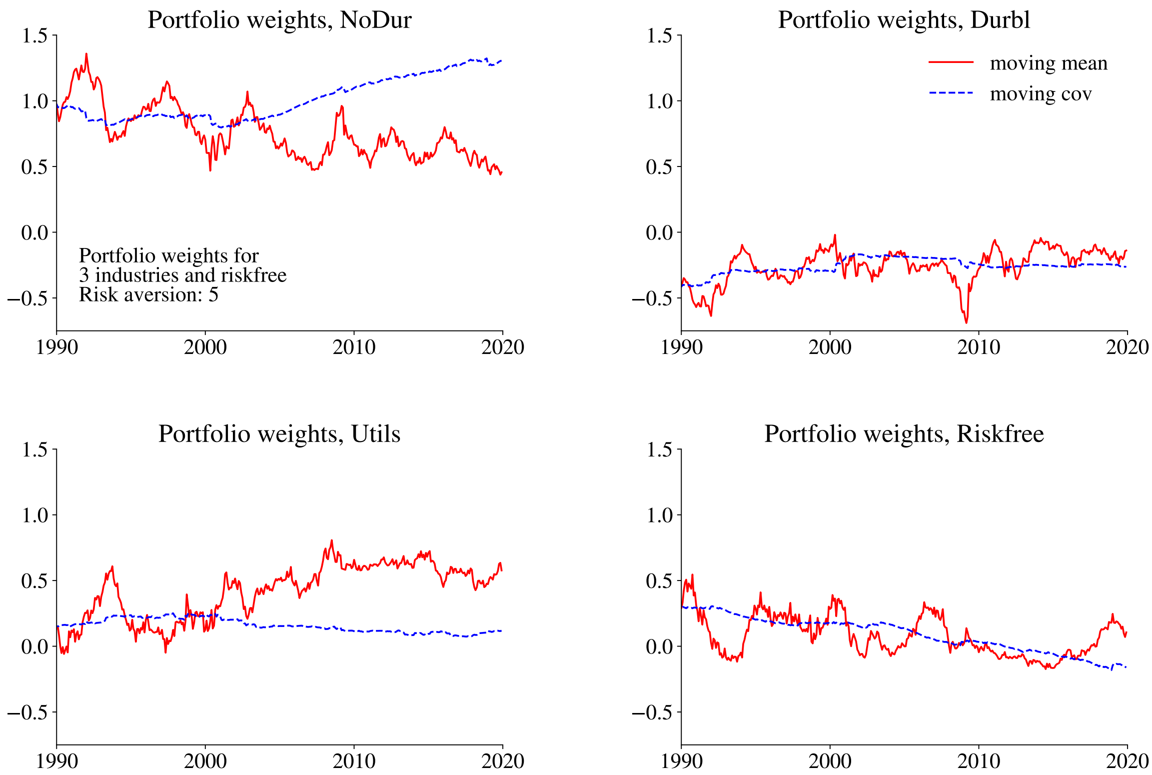


Figure 5.8: Dynamically updated portfolio weights, T-bill and 3 U.S. industries

The solution is $(x, y) = (2.75, -3/4)$, so only one of the bounds is really binding. See Figure 5.11 for an illustration.

Remark 5.9 *With both lower and upper bounds $a \leq x \leq b$, we instead work with the (unbounded) $v = \text{logit}(\frac{x-a}{b-a})$, where the logit function and its inverse are defined as $\text{logit}(u) = \ln(\frac{u}{1-u})$ and $\text{logit}^{-1}(v) = \frac{1}{1+\exp(-v)}$. (The inverse is also called the logistic function.) We can transform back to x by $x = a + (b-a)\,\text{logit}^{-1}(v)$*

5.4.3 Equality Constraints

Suppose you want an *equality constraint* on the minimization problem, say

$$h_1(\theta) = x + 2y - 3 = 0. \tag{5.20}$$

One way to handle this is to use the constraint to rewrite the loss function (in this case, we would use $x = 3 - 2y$ to replace x in (5.18)). If this is tricky, then we try to find

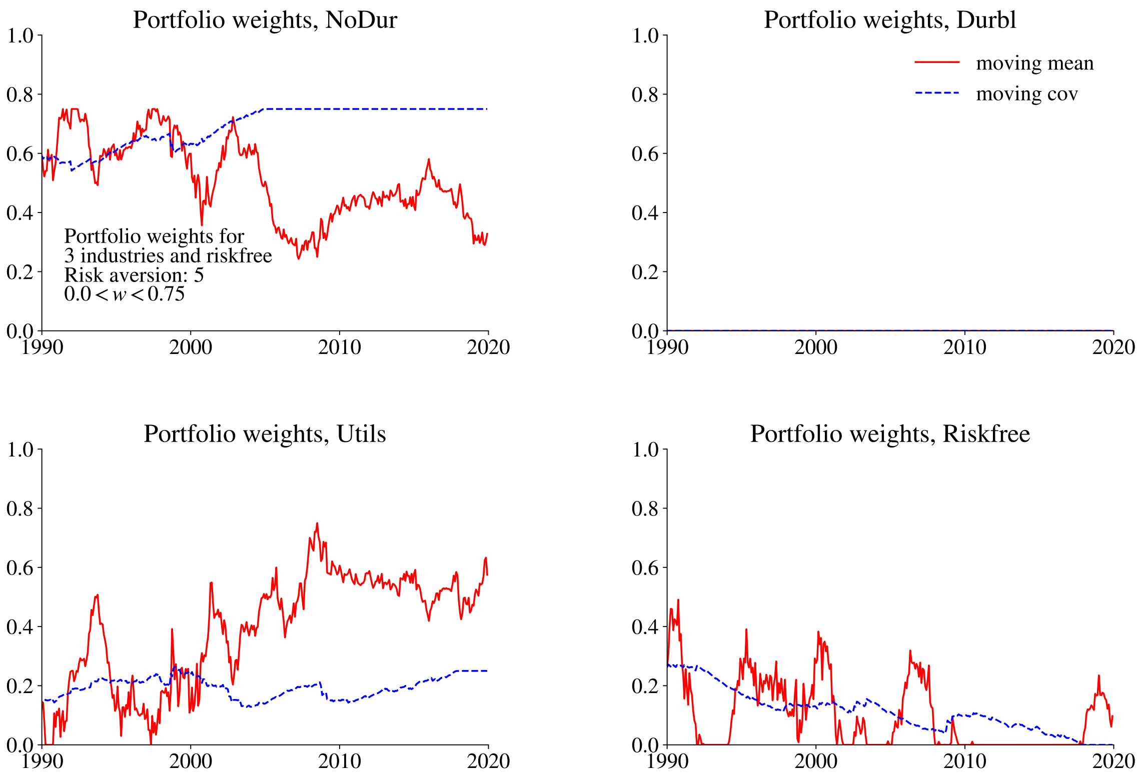


Figure 5.9: Dynamically updated portfolio weights (no short sales), T-bill and 3 U.S. industries

a routine that can handle equality constraints. The short discussion below outlines how these routines work (and also suggests how we could construct such a routine ourselves).

A simply way is to apply a penalty for deviations from the constraint, so the overall loss function becomes

$$f(\theta) + \lambda \textstyle\sum_{i=1}^{p} h_i(\theta)^2, \tag{5.21}$$

where $h_i(\theta)$ is the ith equality constraint. In our example, (5.20) has only one restriction ($p = 1$).

Start by setting $\lambda = 0$ and find the optimal value of θ, and call it θ_1. This is clearly the unconstrained solution. Then, set $\lambda = 1$ (or some other value higher than 0) and redo the optimization (using θ_1 as the starting guess) to get the optimal value θ_2. Now, set $\lambda = 2$ and redo the optimization (using θ_2 as the starting guess). Keep doing this (with higher and higher values of λ) until the solutions do not change much anymore. It is often worthwhile to experiment a bit with the sequence of λ values.

See Figure 5.12 for an example. (The solution should be very close to $(x, y) =$

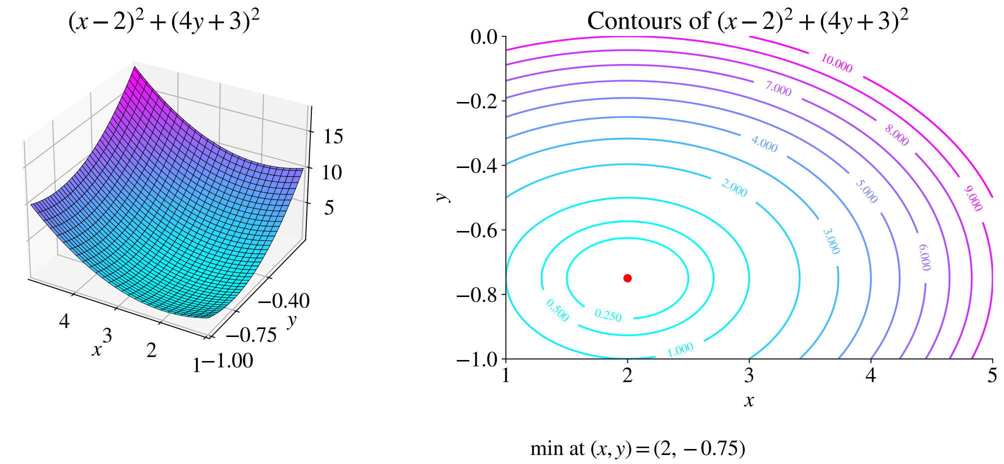


Figure 5.10: Numerical optimization, no restrictions

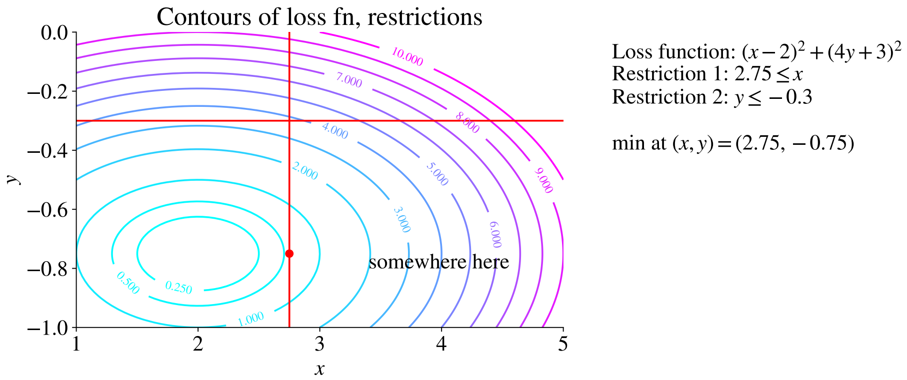


Figure 5.11: Numerical optimization with bounds on the solution

$(4, -1/2)$.)

5.4.4 Inequality Constraints

Instead, we now want to minimize (5.18) under the *inequality constraint* $y \leq -(x-4)^2$.

It is convenient to rewrite all inequality constraints on a common form, and we here choose to write them all on ≤ 0 form, which gives

$$g_1(\theta) = y + (x-4)^2 \leq 0. \tag{5.22}$$

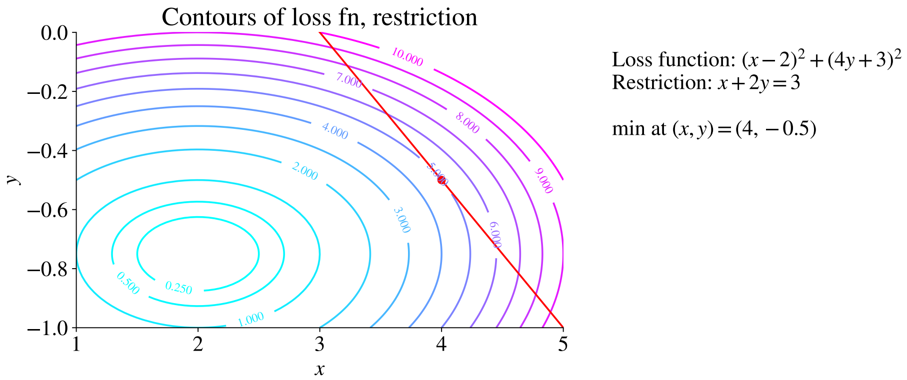


Figure 5.12: Numerical optimization with an equality restriction

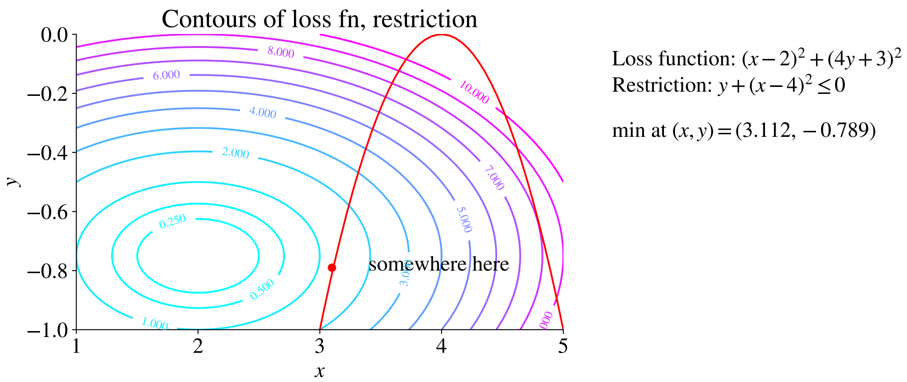


Figure 5.13: Numerical optimization with inequality restriction

Now, we minimise the overall loss function

$$f(\theta) + \rho \textstyle\sum_{j=1}^{q} \max[0, g_j(\theta)]^2, \tag{5.23}$$

where $g_j(\theta)$ is the jth inequality constraint (there is only one in our example). Notice that ρ plays the same role as λ: start by solving for $\rho = 0$, then use that solution as a starting guess for the problem with $\rho = 1$, etc. See Figure 5.13 for an example. (The solution should be close to $(x, y) = (3.1, -0.79)$.)

Finally, we can combine equality and inequality constraints as

$$f(\theta) + \lambda \textstyle\sum_{i=1}^{p} h_i(\theta)^2 + \rho \textstyle\sum_{j=1}^{q} \max[0, g_j(\theta)]^2. \tag{5.24}$$

Chapter 6

CAPM

Reference: Elton, Gruber, Brown, and Goetzmann (2014) 13 and 15
Additional references: Danthine and Donaldson (2005) 7
More advanced material is denoted by a star (*). It is not required reading.

6.1 Beta Representation of Expected Returns

6.1.1 Beta Representation: Definition

If we assume that investors have mean-variance preferences, then it is straightforward to calculate the tangency portfolio. In addition, if we assume that all investors have the same beliefs about expected returns and the covariance matrix (this is restrictive, but can be relaxed—at the cost of making the algebra a lot messier), then the tangency portfolio is the same for all investors. In this setting, a key result (see below for a proof) is that, for any asset, the expected excess return ($\mathrm{E}\, R_i^e$) is linearly related to the expected excess return on the tangency portfolio (μ_T^e) according to

$$\mathrm{E}\, R_i^e = \beta_i \mu_T^e \text{ or} \tag{6.1}$$

$$\mathrm{E}\, R_i = R_f + \beta_i \mu_T^e \tag{6.2}$$

$$\text{where } \beta_i = \sigma_{iT} / \sigma_T^2. \tag{6.3}$$

This result follows directly from manipulating the definition of the tangency portfolio (see below for a proof). Plotting $\mathrm{E}\, R_i^e$ or $\mathrm{E}\, R_i$ against β_i gives the *security market line*, see Figure 6.1. Notice that β_i is the slope from regressing R_i on R_T (using OLS).

Example 6.1 *(Calculating β_i from the covariance matrix) The traditional way of estimating β_i is to run a regression. However, if we know the variance-covariance ma-*

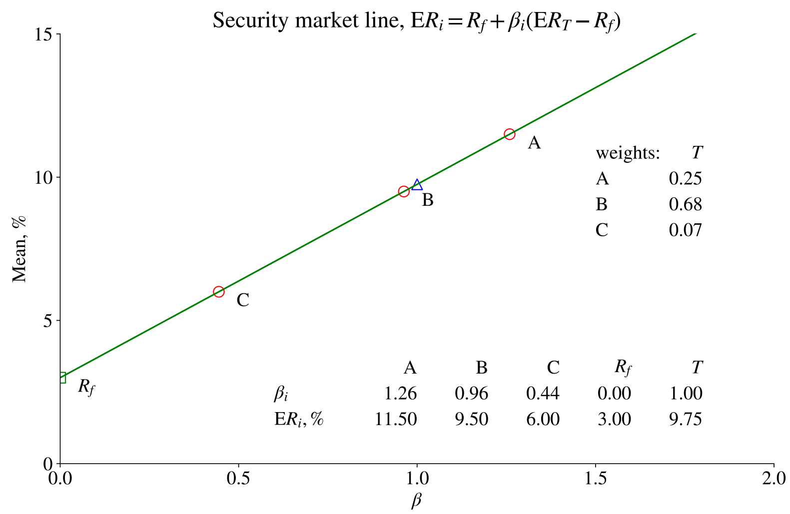


Figure 6.1: Security market line. The properties of the investable assets (A, B, and C) are shown in Table 11.1.

trix Σ of the investable assets, then we can also use the fact that $\beta_i = \sigma_{iT}/\sigma_T^2$ where $\sigma_{iT} = w_i' \Sigma w_T$. Using the asset price characteristics in Table (11.1), together with the weights of the tangency portfolio gives the β values in Figure 6.1.

	μ, %	Σ, bp		
		A	B	C
A	11.5	166	34	58
B	9.5	34	64	4
C	6.0	58	4	100

Table 6.1: Characteristics of the assets in the MV examples. Notice that μ, % is the expected return in % (that is, $\times 100$) and Σ, bp is the covariance matrix in basis points (that is, $\times 100^2$).

Example 6.2 *(Effect of β) With $R_f = 3\%$ and $\mu_T = 9.75\%$ (so $\mu_T^e = 6.75\%$) we get*

β_i	$\mathrm{E}\,R_i$	*Comment*
0.44	6.0%	*baseline case*
1.5	13.12%	*high β, high return*
1	9.75%	*same risk as market*
0	3%	*no risk*
−0.5	−0.38%	*the opposite of risk*

Remark 6.3 *(β of a portfolio) The portfolio with return $R_p = w'R + (1 - \mathbf{1}'w)R_f$ has the beta $\beta_p = w'\beta$. (This follows directly from $\mathrm{Cov}(\Sigma_{i=1}^n w_i R_i, R_T) = \Sigma_{i=1}^n w_i \,\mathrm{Cov}(R_i, R_T)$ and that $\beta_p = \sigma_{pT}/\sigma_T^2$.)*

Remark 6.4 *(Getting the β you want) If you want a portfolio with a specific β (for instance, 1.2), this is easily done by investing β in the tagency portfolio and $1 - \beta$ (here -0.2) in the riskfree. (This follows directly from the "β of a portfolio" and the fact that the tangency portfolio has a beta of one.)*

Example 6.5 *Let $(\beta_1, \beta_2) = (1.2, 0.8)$. The portfolio return $R_p = 0.6R_1 + 0.4R_2$ has the beta $\beta_p = 0.6 \times 1.2 + 0.4 \times 0.8 = 1.04$.*

Proof. (of (6.1)) Let (w_1, w_2) be the tangency portfolio. First, notice that the covariance of asset i (1 or 2) and the tangency portfolio is

$$\sigma_{iT} = \mathrm{Cov}\left(R_i, w_1 R_1 + w_2 R_2\right) = w_1\sigma_{i1} + w_2\sigma_{i2}.$$

This holds also for $i = T$, so

$$\sigma_T^2 = \mathrm{Cov}\left(w_1 R_1 + w_2 R_2, R_T\right) = w_1\sigma_{1T} + w_2\sigma_{2T}.$$

Second, recall the first order conditions for optimal portfolio choice for the investor with risk aversion k_T such that he/she holds the tangency portfolio

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \mu_1^e \\ \mu_2^e \end{bmatrix} - k_T \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}.$$

Use the result on the covariances to rewrite the first order conditions as

$$\begin{bmatrix} \mu_1^e \\ \mu_2^e \end{bmatrix} = \begin{bmatrix} \sigma_{1T} \\ \sigma_{2T} \end{bmatrix} k_T. \tag{*}$$

Third, rewrite σ_T^2 by using (*)

$$\sigma_T^2 = \left(w_1\mu_1^e + w_2\mu_2^e\right)/k_T = \mu_T^e/k_T.$$

Fourth, solve for $k_T = \mu_T^e/\sigma_T^2$ and use in (*) to get

$$\begin{bmatrix} \mu_1^e \\ \mu_2^e \end{bmatrix} = \begin{bmatrix} \sigma_{1T}/\sigma_T^2 \\ \sigma_{2T}/\sigma_T^2 \end{bmatrix} \mu_T^e,$$

which is (6.1). ■

6.1.2 Beta Representation: The Basic Mechanism

CAPM implies that an asset has the same average return as a MV efficient portfolio with the same systematic risk—although it may have a much higher volatility. See Figure 6.2 for an illustration. To formalise this, consider a CAPM regression

$$R_i^e = \alpha_i + \beta_i R_T^e + \varepsilon_i, \tag{6.4}$$

which has the usual property that the residual is uncorrelated with the regressor. We can therefore write the variance as

$$\sigma_i^2 = \beta_i^2\sigma_T^2 + \sigma_\varepsilon^2. \tag{6.5}$$

This says that the variance of return i has two components: *systematic risk* (the comovement of R_i with R_T) and *idiosyncratic noise* (the movements of ε_i).

Now, consider a portfolio on the capital market line (an optimal portfolio), $R_{opt} = vR_T + (1-v)R_f$. Its expected excess return and variance are

$$\mu_{opt}^e = v\mu_T \tag{6.6}$$

$$\sigma_{opt}^2 = v^2\sigma_T^2. \tag{6.7}$$

We now find the optimal portfolio with the same systematic risk as asset i: set $\sigma_i^2 = \sigma_{opt}^2$ and use (6.5)–(6.7)

$$v^2\sigma_T^2 = \beta_i^2\sigma_T^2, \text{ so} \tag{6.8}$$

$$v = \beta_i, \tag{6.9}$$

provided v and β_i have the same signs. The second equation says that the optimal portfolio (with the appropriate systematic risk) has $v = \beta_i$. If we now requires that the asset i

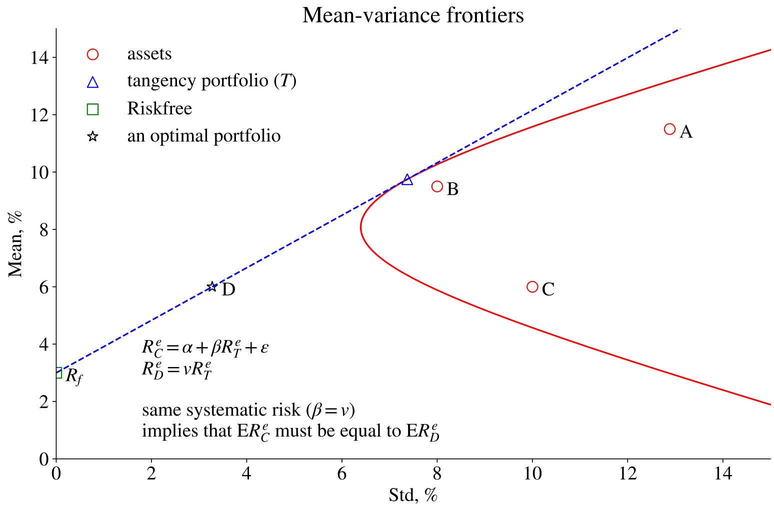


Figure 6.2: Mean-variance frontier and expected returns

has the same average (excess) return as this optimal portfolio, then that means

$$\mathrm{E}\, R^e_i = \beta_i \mu^e_T, \tag{6.10}$$

which is CAPM.

Example 6.6 *In Figure 6.2, we want to understand the mean return (vertical location) of asset C (taking its volatility and β as given). We notice that C has the same systematic risk as the efficient portfolio D. According to CAPM, C must then have the same average return as D.*

Example 6.7 *(Creating a portfolio with $\beta_p = 0.44$) We can create a portfolio with $\beta = 0.44$ by investing 0.44 into the tangency portfolio and 0.56 in the riskfree.*

Remark 6.8 *(Zero beta portfolio*) Suppose we create a portfolio by buying $1/\beta_i$ (in value terms) of asset i and $-1/\beta_j$ of asset j and keeping the rest in the riskfree asset. The beta of this portfolio is*

$$\beta_p = \frac{\beta_i}{\beta_i} - \frac{\beta_j}{\beta_j} + (1 - 1/\beta_i - 1/\beta_j)0 = 0,$$

since the beta of a portfolio is the portfolio of the betas, and the riskfree has a zero beta. According to (6.1), the expected return of this portfolio should be zero—since it has no systematic risk.

Remark 6.9 *(Why is Risk = β? Alternative version*) Start by investing 100% in the tangency 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*

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

The expected portfolio return is

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

and the portfolio variance is

$$\mathrm{Var}(R_p) = \sigma_T^2 + \underbrace{\delta^2\sigma_i^2 + 2\delta\sigma_{iT}}_{\text{incremental risk, but } \delta^2\sigma_i^2 \approx 0}.$$

(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

$$\frac{\text{incremental risk premium}}{\text{incremental risk}} = \frac{\mathrm{E}\, R_i^e}{2\sigma_{iT}}.$$

In equilibrium, the marginal compensation for more risk must be equal across assets—or else it would be optimal to deviate from the tangency portfolio by going long in assets with a favourable ratio—and vice versa. That is, we must have

$$\frac{\mathrm{E}\, R_i^e}{2\sigma_{iT}} = \frac{\mathrm{E}\, R_j^e}{2\sigma_{jT}} = \dots = \frac{\mathrm{E}\, R_T^e}{2\sigma_T^2},$$

since $\mathrm{Cov}\,(R_T, R_T) = \sigma_T^2$. *This can be rearranged as the CAPM expression.*

6.1.3 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 matrix). They will then all mix the tangency portfolio with the riskfree—but possibly in different proportions due to different risk aversions.

In equilibrium, net supply of the riskfree assets is zero (lending = borrowing), so *the average investor must hold the tangency portfolio and no riskfree assets*. Therefore, *the tangency portfolio must be the market portfolio*.

6.1.4 Properties of the Market Portfolio

It is straightforward to show that the market risk premium (expected excess return) is proportional to the market volatility

$$\mathrm{E}\, R_m^e = k_m \sigma_m^2, \tag{6.11}$$

where we used the subscript m to indicate that this is the market portfolio (which equals the tangency portfolio).

Proof. (of (6.11)) Recall the first order conditions for optimal portfolio choice for the investor with risk aversion k_T such that he/she holds the tangency portfolio

$$\begin{bmatrix} \mu_1^e \\ \mu_2^e \end{bmatrix} = k_T \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}.$$

Premultiply both sides by $\begin{bmatrix} w_1 & w_2 \end{bmatrix}$ and set $m = T$ to get (6.11). ■

Combining (6.11) with the beta representation (6.1) (and setting $m = T$) gives

$$\begin{aligned} \mathrm{E}\, R_i^e &= \beta_i \mu_m^e \\ &= \beta_i k_m \sigma_m^2. \end{aligned} \tag{6.12}$$

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 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 (6.12) 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.

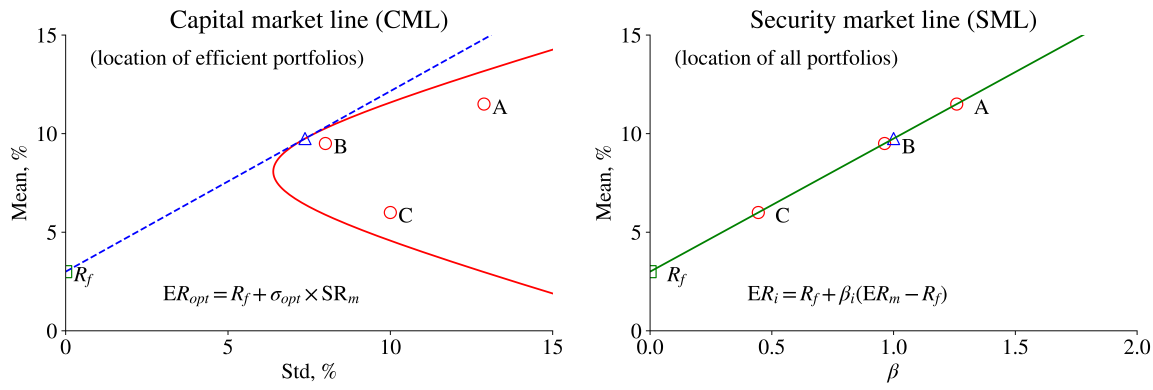


Figure 6.3: CML and SML

6.1.5 Summarizing MV and CAPM: CML and SML

According to MV analysis, the average return on all optimal (effective) portfolios (denoted opt) obey

$$\mathrm{E}\,R_{opt} = R_f + \sigma_{opt} SR_m. \tag{6.13}$$

The plot of $\mathrm{E}\,R_{opt}$ against σ_{opt} is called the *capital market line*. See Figure 6.3 for an example.

According to CAPM, the average return on all portfolios (optimal or not), obey the beta representation (6.2)

$$\mathrm{E}\,R_i = R_f + \beta_i\,\mathrm{E}\,R^e_m. \tag{6.14}$$

The plot of $\mathrm{E}\,R_i$ against β_i (for different assets, i) is called the *security market line*. See Figure 6.3 for an example.

6.1.6 Using CAPM to Measure Asset Performance

Consider a portfolio q that we want to evaluate—and write its average excess return as

$$\mathrm{E}\,R^e_q = \alpha + \beta_q \mu^e_m. \tag{6.15}$$

We can easily *replicate* the $\beta_q\,\mathrm{E}\,R^e_m$ part by constructing a portfolio

$$R_p = \beta_q R_m + (1 - \beta_q) R_f, \text{ so} \tag{6.16}$$

$$\mathrm{E}\,R^e_p = \beta_q \mu^e_m. \tag{6.17}$$

Portfolio q will have both systematic and idiosyncratic volatility (see (6.5)), while portfolio p has the same systematic volatility but no idiosyncratic volatility. Hence, portfolio q must be at least as volatile as p.

However, portfolio q may still perform better (bring "value added") than p if $\alpha > 0$. This motivates why α can be used to *measure performance*. However, this assumes that *(i)* CAPM is (mostly) valid; *(b)* but some assets (like q) may deviate from CAPM.

Example 6.10 *(α as a performance measure) Suppose* $\mathrm{E}\,R_q^e = 10\%$*,* $\beta_p = 1.2$ *and* $\mu_m^e = 9\%$*. Construct a portfolio p so that* $R_p = 1.2R_m - 0.2R_f$ *(buy 1.2 of an ETF on the market index and borrow 0.2). Notice that* $\mathrm{E}\,R_p^e = 1.2 \times 9\% = 10.8\%$*. This means that asset q has* $\alpha = -0.8\%$*: the replicating portfolio has a higher return (and it can be shown that it will also have a lower volatility).*

6.1.7 Back to Prices (The Gordon Model)*

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

$$1 + R_{t+1} = \frac{D_{t+1} + P_{t+1}}{P_t}, \tag{6.18}$$

where P_t is the asset price and D_{t+1} is the dividend it gives at the beginning of the next period. If we assume that expected returns are constant across time (denoted R, for instance 10%) and that dividends are expected to grow at the rate g (for instance, 2%), then it is straightforward to show that the asset price is

$$P_t = \mathrm{E}_t\, D_{t+1} \sum_{s=1}^{\infty} \frac{(1+g)^{s-1}}{(1+R)^s} = \frac{\mathrm{E}_t\, D_{t+1}}{R-g}. \tag{6.19}$$

Clearly, higher (expected) dividends and/or a higher growth rate increases the asset price. In addition, a lower expected ("required") *future return* also increases *today's asset price*.

In CAPM, a lower expected return could be driven by a lower beta or by a lower riskfree rate. One way of interpreting this is as follows. If an asset (suddenly) gets a lower beta, that means that it has less systematic risk than before. It is therefore more useful in portfolio formation (more diversification benefits) and becomes more demanded—so the price level increases. With a higher price level, the dividend yield is lower, which contributes to a lower return (recall the return is the dividend yield plus the capital gains yield).

6.2 Testing CAPM

Reference: Elton, Gruber, Brown, and Goetzmann (2014) 15

Let $R^e_{it} = R_{it} - R_{ft}$ be the excess return on asset i in excess over the riskfree asset in period t, and let R^e_{mt} be the excess return on the market portfolio in the same period. (The time subscripts are written out to highlight that we use time series data to estimate and test the regression coefficients.) The basic implication of CAPM 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

$$\mathrm{E}\, R^e_{it} = \beta_i \,\mathrm{E}\, R^e_{mt}, \text{ where } \beta_i = \sigma_{im}/\sigma^2_m. \tag{6.20}$$

Consider the regression

$$\begin{aligned} R^e_{it} &= \alpha_i + \beta_i R^e_{mt} + \varepsilon_{it}, \text{ where} \\ \mathrm{E}\, \varepsilon_{it} &= 0 \text{ and } \mathrm{Cov}(R^e_{mt}, \varepsilon_{it}) = 0. \end{aligned} \tag{6.21}$$

The two last conditions are automatically imposed by LS. Take expectations of the regression (assuming we know the coefficients) to get

$$\mathrm{E}\, R^e_{it} = \alpha_i + \beta_i \,\mathrm{E}\, R^e_{mt}. \tag{6.22}$$

Notice that the LS estimate of β_i in (6.21) is the sample analogue to β_i in (6.20), since LS estimates a slope coefficient as the covariance of the dependent variable and the regressor, divided by the variance of the regressor. 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 (6.20) 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}. That is, we test if the market portfolio is the correct "pricing factor" of all the test assets. This is the perspective of CAPM tests.

The 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

$$\frac{\hat{\alpha}_i}{\text{Std}(\hat{\alpha}_i)} \xrightarrow{d} N(0,1) \text{ under } H_0 : \alpha_i = 0. \tag{6.23}$$

In this expression, $\hat{\alpha}_i$ is the estimate of the intercept in (6.21) and $\text{Std}(\hat{\alpha}_i)$ its standard deviation (for instance, from the usual OLS results). 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. We typically reject the null hypothesis ($\alpha_i = 0$) when the the t-statistic is very negative or very positive (for instance, lower than -1.96 or higher than 1.96).

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.

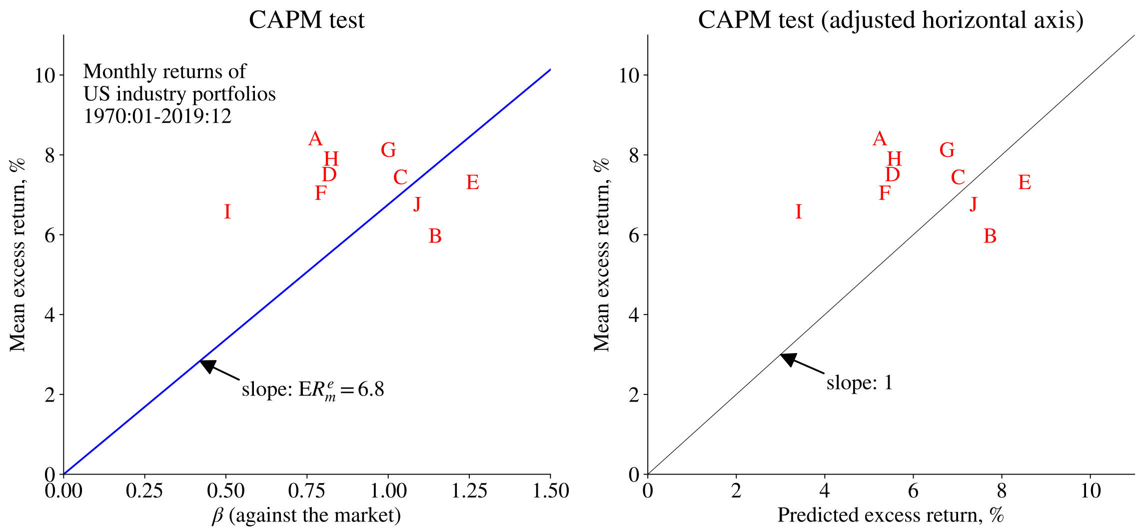
The empirical results from such tests vary with the test assets used. For US portfolios, CAPM seems to work reasonably well for some types of portfolios (for instance, portfolios based on firm size or industry), but much worse for other types of portfolios (for instance, portfolios based on firm dividend yield or book value/market value ratio). Figure 6.4 shows some results for US industry portfolios.

6.2.1 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 judgment on. For a more formal approach, a SURE approach is useful. Alternatively, we can apply a Bonferroni correction of the individual t-stats: reject CAPM at the 5% significance level only if the largest t-stat (in absolute terms) exceeds the critical value at the $0.05/n$ significance level. For instance, with $n = 25$, the critical value from a standard normal distribution would be 3.09 instead of 1.96.

A quite different approach to study a cross-section of assets is to first perform a CAPM



	α	t-stat	σ
A (NoDur)	3.12	2.55	8.59
B (Durbl)	-1.82	-0.97	13.18
C (Manuf)	0.38	0.43	6.16
D (Enrgy)	1.93	0.92	14.83
E (HiTec)	-1.25	-0.77	11.37
F (Telcm)	1.65	1.08	10.66
G (Shops)	1.31	1.01	9.10
H (Hlth)	2.28	1.46	11.01
I (Utils)	3.12	1.90	11.53
J (Other)	-0.66	-0.69	6.72

CAPM: $R_i^e = \alpha_i + \beta_i R_m^e + e_i$
Predicted excess return: $\beta_i \mathrm{E}R_m^e$

α and σ (std of residual) are in annualized %

p-val for testing if all $\alpha_i = 0$: 0.171
10% crit. value (Bonferroni): 2.58

Figure 6.4: CAPM regressions on US industry indices

regression (6.21) and then the following cross-sectional regression

$$\bar{R}_i^e = \gamma + \lambda \hat{\beta}_i + u_i, \tag{6.24}$$

where $\bar{R}_i^e$ 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 (6.24) to have a high R^2—since it should be unity in a very large sample (if CAPM holds).

6.2.2 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:

$$\begin{bmatrix} \text{small size, low B/M} & \dots & \dots & \dots & \text{small size, high B/M} \\ \vdots & \ddots & & & \\ \vdots & & \ddots & & \\ \vdots & & & \ddots & \\ \text{large size, low B/M} & & & & \text{large size, high B/M} \end{bmatrix} \tag{6.25}$$

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). 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. 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. Figure 6.4 shows some results for US industry portfolios and Figures 6.5–6.6 for US size/book-to-market portfolios.

In Figure 6.4, the results are presented in two different ways:

$$\begin{array}{lll} & \underline{\text{horizontal axis}} & \underline{\text{vertical axis}} \\ 1: & \beta_i & \sum_{t=1}^{T} R^e_i / T \\ 2: & \beta_i \sum_{t=1}^{T} R^e_m / T & \sum_{t=1}^{T} R^e_i / T \end{array} \tag{6.26}$$

In the first approach, CAPM 6.20 says that all data points (different assets, i) should cluster around a straight line with a slope equal to the average market excess return,

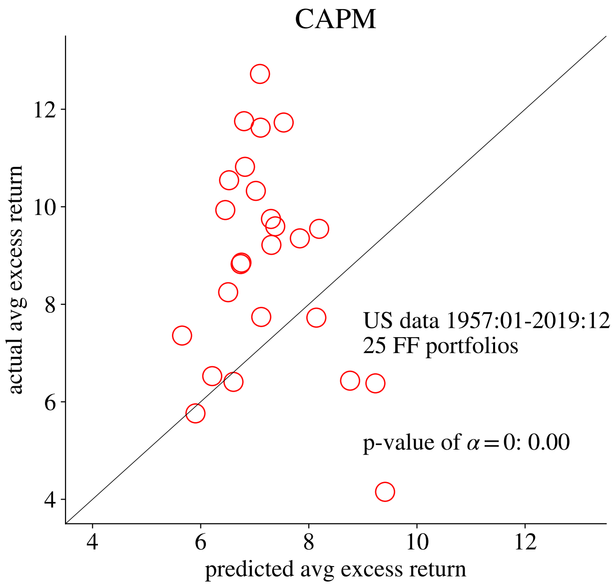


Figure 6.5: CAPM, FF portfolios

$\sum_{t=1}^{T} R_m^e / T$. In the second approach, CAPM says that all data points should cluster around a 45-degree line. In either case, the vertical distance to the line is α_i (which should be zero according to CAPM).

6.2.3 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 bad funds).

Example 6.11 *(Steadman's funds*) "How can a fund be this bad?" (NYT, 1991) (the four Steadman funds rank among the six worst performers of the 244 stock funds tracked by Lipper Analytical Services for the 15 years that ended on Oct. 31. The Oceanographic Fund comes in at No. 243 and Steadman American Industry Fund, No. 244); "Steadman's creature just won't die" (Forbes, 1999); "Those awful Steadman's funds returning under a new name" (Baltimore Sun, 2002).*

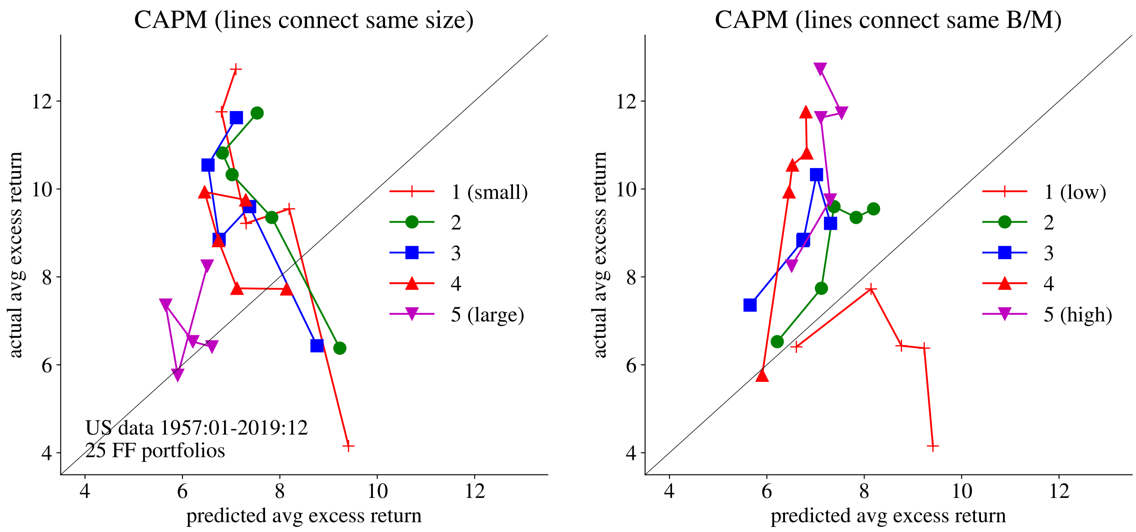


Figure 6.6: CAPM, FF portfolios

6.3 Appendix: Statistical Tables*

Tables 6.2 shows critical values for t-distributions (and the standard normal), while 6.3 shows critical values for chi-square distributions.

n	Significance level		
	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 6.2: Critical values (two-sided test) of t distribution (different degrees of freedom) and normal distribution.

n	Significance level		
	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 6.3: Critical values of chisquare distribution (different degrees of freedom, n).

Chapter 7

Foreign Exchange

Reference: Elton, Gruber, Brown, and Goetzmann (2014) 12, Sercu (2009).
More advanced material is denoted by a star (*). It is not required reading.

7.1 Exchange Rate Basics

7.1.1 Exchange Rate Quotation

An exchange rate is the price of one currency in terms of another currency. There are clearly *two ways of quoting* this price. For instance, CHF 1 may cost EUR 0.8333 (sometimes written as "CHF 1 = EUR 0.8333") which clearly means that EUR 1 would cost CHF 1.2 ("EUR 1 = CHF 1.2"). There is a reasonably established set of quotations on the interbank market, but in other settings either type of quotation is possible: always double check.

Remark 7.1 *(The meaning of CHF/EUR) These lecture notes follow the convention that CHF/EUR (or $S^{CHF/EUR}$) denotes how many CHF you have to pay for each EUR (CHF/EUR is how many CHF* per *EUR), for instance, $S^{CHF/EUR} = 1.2$. Clearly, $S^{x/y} = 1/S^{y/x}$, for instance, $S^{EUR/CHF} = 0.8333$. (In contrast, the interbank FX market often use EUR/CHF EURCHF to denote the same thing, that is, how many CHF you get for one EUR.)*

Remark 7.2 *(Currency codes, according to ISO 4217) USD, EUR, JPY, GBP, AUD, CAD, CHF, CNY (Chinese yuan), SEK (Swedish krona), MXN (Mexican peso).*

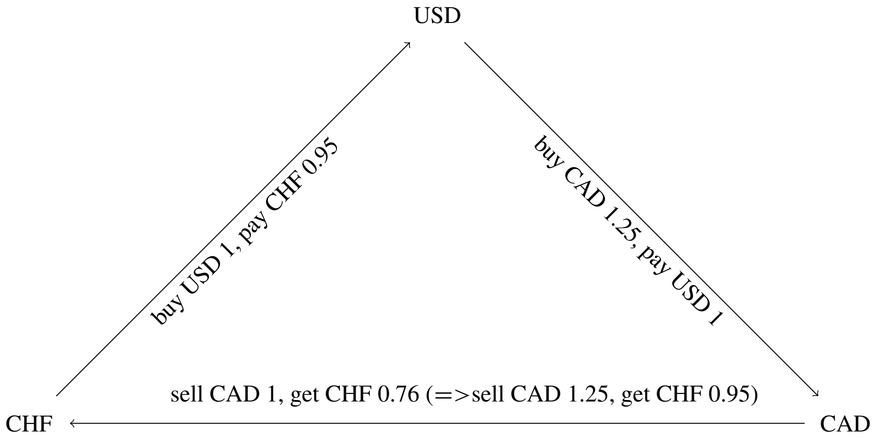


Figure 7.1: Cross-rates

7.1.2 Cross Rates

Exchange rate across "smaller" currencies are often established indirectly and are therefore called "cross rates": as a combination of two trades. For instance, suppose you own CHF and want to buy CAD (Canadian dollars). It may well be that this involves two trades: use the CHF to buy USD and then use the USD to buy CAD. (Even 15 years after the collapse of the Bretton-Woods system in the early 1970s almost all currency trades went via the USD. Since then there are much more direct trades.)

Example 7.3 *(The implicit trade in a cross rate) (a) Buy one USD, costs* 0.95 *CHF; (b) use the one USD to buy* 1.25 *CAD; (c) in total you have paid* 0.95 *CHF and got* 1.25 *CAD. Therefore, the implied price (in CHF) per AUD is* $0.95/1.25 \approx 0.76$*. (You can memorize this as "CHF/USD×USD/CAD=CHF/CAD") See Figure 7.1 for an illustration.*

Remark 7.4 *(The implicit trade in a cross rate, using* $S^{x/y}$ *notation) In the previous example,* $S^{CHF/USD} = 0.95$ *and* $S^{USD/CAD} = 1/1.25$*, so*
$S^{CHF/USD} = S^{CHF/USD}S^{USD/CAD} = 0.95/1.25 \approx 0.76$*. In general, cross rates mean that* $S^{x/y} = S^{x/z}S^{z/y}$ *(or equivalently* $S^{x/z}/S^{y/z}$ *and* $S^{z/y}/S^{z/x}$*).*

If there is a way to trade without going through another currency (and there typically is), then the price on this market should be the same (or at least very close to) the cross rate. If not, there would be an arbitrage opportunity.

$t-1$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ t

(a) buy 1 foreign currency unit: pay S_{t-1}
(b) lend on foreign money market

(a) collect $1 + R_f^*$
(b) sell foreign currency: get $(1 + R_f^*)S_t$

Return: $R_t = (1+R_f^*)S_t/S_{t-1}-1$

UIP: $\mathrm{E}_{t-1}R_t = R_f \Leftrightarrow$
$\mathrm{E}_{t-1}S_t/S_{t-1} = (1+R_f)/(1+R_f^*)$

Figure 7.2: Return on currency investment

7.2 Investing in Foreign Currency

7.2.1 The Return from Holding Currency

Investing in a foreign currency typically means that you buy that currency, lend on the foreign money market and eventually buy back domestic currency. To define the return, let the simplified notation S_t be today's price (measured in domestic currency) of one unit of foreign currency (the "asset"), that is, let S_t be short hand notation for $S_t^{\text{home/foreign}}$. Also, let R_{ft}^* be the foreign riskfree rate between $t-1$ and t. The *return* (measured in domestic currency) is then

$$R_t = (1 + R_{ft}^*)\frac{S_t}{S_{t-1}} - 1. \tag{7.1}$$

Remark 7.5 *(*Details of the currency return R_t) In $t-1$ invest S_{t-1} (of domestic currency) to buy one unit of foreign currency and lend it. After one period you have $1 + R_f^*$ units of foreign currency, which buys $(1 + R_f^*)S_t$ units of domestic currency (this is the payoff). The gross return is payoff/investment, which is $(1 + R_f^*)S_t/S_{t-1}$.*

The return of the foreign investment *in excess of the domestic riskfree rate* is then

$$R_t^e = (1 + R_{ft}^*)\frac{S_t}{S_{t-1}} - (1 + R_{ft}). \tag{7.2}$$

Clearly, an appreciation of the foreign currency (or a depreciation of the domestic currency) as well as a higher foreign riskfree rate are positive for the return, while a higher domestic riskfree rate ("financing cost") is negative. See Figure 7.3.

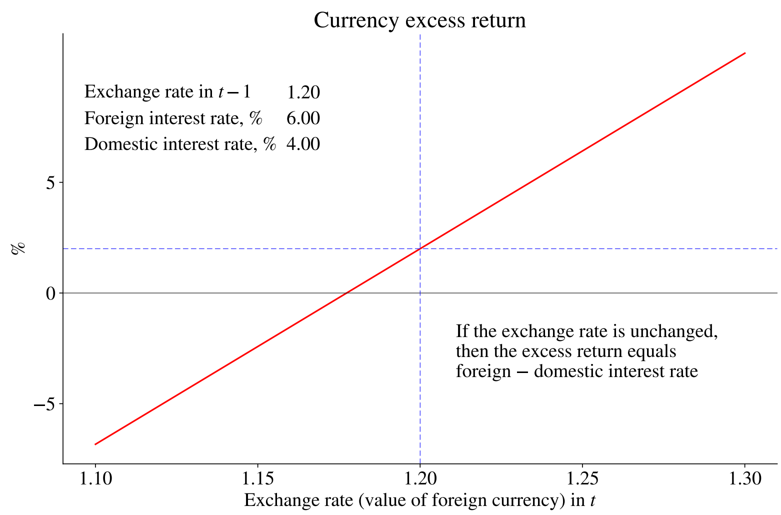


Figure 7.3: Illustration of currency excess returns as a function of the realized exchange rate

Example 7.6 *With* $(S_{t-1}, S_t, R_f^*, R_f) = (1.20, 1.25, 0.06, 0.04)$

$$R_t^e = (1 + 0.06)\frac{1.25}{1.20} - (1 + 0.04) = 0.064.$$

With $S_t = 1.20$ *the excess return is* $0.06 - 0.04 = 0.02$*. Instead with* $S_t = 1.177$ *the excess return is close to zero*

$$R_t^e = (1 + 0.06)\frac{1.177}{1.20} - (1 + 0.04) \approx 0.$$

Remark 7.7 *(*Log FX returns) Let* r_t *be the log return,* $\ln(1 + R_t)$*. From (7.1), it can be written* $r_t = r_{ft}^* + \Delta s_t$*,where* r_{ft}^* *is the log foreign gross riskfree rate,* $\ln(1 + R_{ft}^*)$*, and* Δs_t *is the relative change of the exchange rate,* $\ln S_t / S_{t-1}$*. Subtract* $r_{ft} = \ln(1 + R_{ft})$ *to get the excess log return* $r_t^e = \Delta s_t + r_{ft}^* - r_{ft}$*.*

Remark 7.8 *(*Alternative exchange rate quotation) If you instead work with exchange rate quotes that use the number of foreign currency units needed to buy one domestic currency unit,* $\tilde{S}$*, then replace* S *by* $1/\tilde{S}$ *in the previous equations.*

In practice, *riskfree returns are from zero-coupon bonds* (bills). We can rewrite $1+R_{ft}$ in terms of an interest as

$$1 + R_{ft} = (1 + Y_{t-1})^m, \tag{7.3}$$

where Y_{t-1} is an effective *annualized interest rate* determined in $t-1$ and m is the fraction of a year between date $t-1$ and t (for instance, $m = 1/12$ for a month and $m = 2$ for two years). Notice that the interest rate is dated $t-1$, since we know already then how much we earn on the bond between $t-1$ and t. For the foreign market, we have $1 + R^*_{ft} = (1 + Y^*_{t-1})^m$. Using this in (7.2) gives the excess return on the foreign investment as

$$R^e_t = (1 + Y^*_{t-1})^m \frac{S_t}{S_{t-1}} - (1 + Y_{t-1})^m. \tag{7.4}$$

Remark 7.9 *(Log FX returns*) (7.3) can be used to rewrite the log return in Remark 7.7 as $r^e_t = \Delta s_t + m(y^*_{t-1} - y_{t-1})$, where $y = \ln(1 + Y)$.*

7.2.2 Covered Interest Rate Parity

To avoid arbitrage opportunities, the forward price (F_{t-1}) set in $t-1$ for delivery of one unit of foreign asset in t must obey

$$F_{t-1} = \frac{1 + R_{ft}}{1 + R^*_{ft}} S_{t-1}. \tag{7.5}$$

This is an application of the spot-forward parity, which for the FX market is often called covered interest rate parity (CIP). In practice, there are (albeit small) deviations from CIP.

Example 7.10 *Using the same numbers as in Example 7.6 we get*

$$F_{t-1} = \frac{1 + 0.04}{1 + 0.06} \times 1.20 \approx 1.177.$$

Proof. (of (7.5)) Replace the risky strategy in (7.2) by "locking in" the FX rate with a forward contract (replace S_t by F_{t-1}) to get $(1 + R^*_{ft})\frac{F_t}{S_{t-1}} - (1 + R_{ft})$. This riskfree return must be zero, or else arbitrageurs step in. Rearrange as (7.5). ■

We can use CIP to rewrite the excess return (7.2) as

$$R^e_t = (1 + R_{ft})\frac{S_t}{F_{t-1}} - (1 + R_{ft}). \tag{7.6}$$

Notice, this is the excess return on the same investment strategy as in (7.2), that is, buy foreign currency in $t-1$ and sell it in t. However, it could also be interpreted as the excess return based on entering a forward contract in $t-1$ and selling the foreign currency in t (see the remark below).

Remark 7.11 *(*Log FX returns) Take logs of (7.5), rearrange and use in the excess log return in Remark 7.7 to get* $r_t^e = \Delta s_t - (f_{t-1} - s_{t-1}) = s_t - f_{t-1}$*. Also, the interest rate differential can be written* $r_{ft}^* - r_{ft} = s_{t-1} - f_{t-1}$*, where* r_{ft} *is the log return,* $\ln(1 + R_{ft})$*.*

7.2.3 Uncovered Interest Rate Parity

The uncovered interest rate parity (UIP) says that the expected exchange rate ($\mathrm{E}_{t-1} S_t$) must be such that the expected excess return from the currency speculation in (7.2) is zero

$$\mathrm{E}_{t-1} R_t^e = 0. \tag{7.7}$$

This means that investing on the foreign money market (and then changing back to the domestic currency) or the domestic money market have the *same expected returns*—in spite of having very different risks. Notice that this is very different from CIP which only rules arbitrage opportunities and says nothing about risk. A somewhat more flexible form of UIP would add a *constant* risk premium to (7.7).

A zero expected excess return in (7.2) means that we must have

$$\frac{\mathrm{E}_{t-1} S_t}{S_{t-1}} = \frac{1 + R_{ft}}{1 + R_{ft}^*}. \tag{7.8}$$

UIP thus says that the domestic currency is expected to appreciate ($\mathrm{E}_{t-1} S_t / S_{t-1} < 1$) if the domestic interest rate is lower than the foreign. (Recall that a lower S means that you pay less to buy a unit of foreign currency: the domestic currency is worth more.) In this way, the foreign investment gains from the interest rate, but loses from the (expected) exchange rate movement—leaving the (expected) return the same as in the domestic market.

Example 7.12 *(UIP) Using the same number as in Example 7.6, UIP says that*

$$\mathrm{E}_{t-1} S_t = 1.20 \times \frac{1 + 0.04}{1 + 0.06} = 1.177,$$

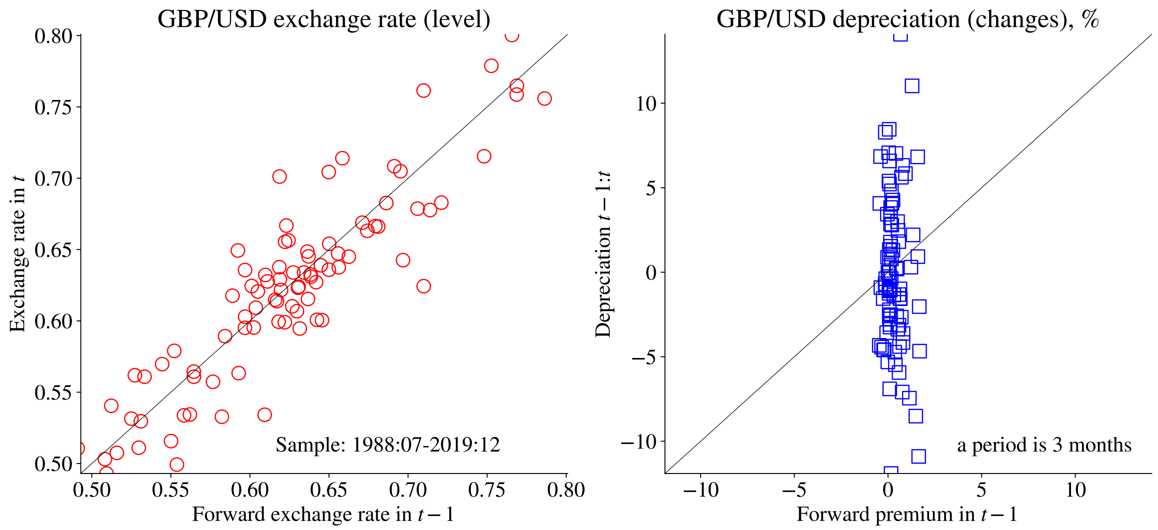
so the domestic currency is expected to appreciate.

Remark 7.13 *(*UIP in terms of interest rates or forwards) Using the definition of the riskfree rate in (7.3) and CIP (7.5), we can rewrite UIP (7.8) as*

$$\frac{\mathrm{E}_{t-1} S_t}{S_{t-1}} = \frac{(1 + Y_{t-1})^m}{(1 + Y_{t-1}^*)^m} = \frac{F_{t-1}}{S_{t-1}}.$$

This says $\mathrm{E}_{t-1} S_t = F_{t-1}$.

Empirical evidence suggests considerable deviations from UIP, which could be explained by either large movements in the risk premia over time or that there have been systematic surprises in historical samples. See Figure 7.4 for an example (it shows how to predict the exchange rate S_t by for instance F_{t-1} and the exchange rate depreciation $S_t/S_{t-1} - 1$ by the forward premium $F_{t-1}/S_{t-1} - 1$).



To predict the exchange rate level, S_t, use:
(a) regression: $a + bF_{t-1}$ from OLS
(b) random walk: S_{t-1}
(c) forward rate: F_{t-1}

To predict the depreciation, $S_t/S_{t-1} - 1$, use:
(a) regression: $a + bF_{t-1}/S_{t-1}$ from OLS
(b) random walk: 0
(c) forward premium: $F_{t-1}/S_{t-1} - 1$

(MSE)	Level	Depreciation
OLS	8.23	20.92
RW	8.31	21.05
Forward	8.61	21.81

Figure 7.4: GBP/USD spot and forward exchange rates

7.2.4 Carry Trade

A common FX strategy is to borrow a low interest rate currency (CHF, JPY?), buy a high interest rate currency (AUD?) and lend on its money market. This called a *carry trade*. It has a positive return if the high interest rate currency depreciates less than suggested by UIP, but clearly also carries the risk of the opposite happening. Empirical evidence suggests that carry trades have generated positive average returns, but are exposed to (intermittent) dramatic losses.

The excess return of a carry trade is given by R_t^e in (7.2). However, a carry trade need not borrow the domestic currency. For instance, a US investor could borrow JPY and lend AUD. Clearly, this strategy would benefit from an appreciation of the AUD and a depreciation of the JPY, as well from a high AUD interest rate and low JPY interest rate.

The remarks below discuss some of the details.

Remark 7.14 *(*Excess return from a carry trade I) Consider the following portfolio in* $t-1$*, you (1) borrow USD, buy and lend AUD; (2) borrow JPY, buy and lend USD. In* t*, you sell/buy the foreign currency for USDs. Using short hand notation for the currencies (U for USD, A for AUD and J for JPY) the excess return (measured in USD) is*

$$R_t^e = (1 + R_{ft}^A)\frac{S_t^{U/A}}{S_{t-1}^{U/A}} - (1 + R_{ft}^J)\frac{S_t^{U/J}}{S_{t-1}^{U/J}},$$

where R_{ft}^A *is the AUD riskfree rate and* R_{ft}^J *is the JPY riskfree rate. Notice that the terms involving the USD riskfree rate cancel, since we are both borrowing and lending USD.*

Remark 7.15 *(*Implementing carry trade with forward contracts) Recall that the gross return on a USD/XXX forward contract is* $S_t^{U/X}/[F_{t-1}^{U/X}/(1 + R_{ft}^U)]$*, where we use the same short hand notation as in Remark 7.14. To see this, notice that the present value of the cost is* $F_{t-1}^{U/X}/(1 + R_{ft}^U)$ *and the payoff is* $S_t^{U/X}$*. Consider a USD/AUD forward and a USD/JPY forward. The difference between the two gross returns is*

$$\frac{S_t^{U/A}}{F_{t-1}^{U/A}/(1 + R_{ft}^U)} - \frac{S_t^{U/J}}{F_{t-1}^{U/J}/(1 + R_{ft}^U)}.$$

This is the excess return earned by going long the USD/AUD forward and short the USD/JPY forward. Also, it is straightforward to use CIP (7.5) to show that this is the same result as in Remark 7.14.

7.3 Currency Risk in Foreign Investments

We now consider an investment into a risky foreign asset. The definition of the return is similar to (7.1), except that we replace the safe foreign return (R_{ft}^*) with a risky foreign return (R_t^*). This means that the foreign investment contributes to the uncertainty about the total return via both the uncertainty in R_t^* and its covariance with the exchange rate movements.

The gross return (measured in domestic currency) of this investment is

$$1 + R_t = (1 + R_t^*)\frac{S_t}{S_{t-1}}. \tag{7.9}$$

Take logs to get

$$r_t = r_t^* + \Delta s_t, \tag{7.10}$$

where r_t^* is the log foreign return, $\ln(1 + R_t^*)$, and Δs_t is the change of the log exchange rate, $\ln(S_t/S_{t-1})$. Notice that our investor gains if *(a)* foreign equity increases in value ($r_t^* > 0$) and *(b)* if the foreign currency increases in value (appreciates) relative to the domestic currency ($\Delta s_t > 0$). See Figures 7.5–7.7 for empirical illustrations.

Example 7.16 *(Investing abroad) Consider a US investor buying British equity in period* $t - 1$

5.5 GBP per British share × 1.6 USD per GBP = 8.8 USD,

and selling in t

5.1 *GBP per British share × 1.9 USD per GBP = 9.69 USD.*

The gross return for the US investor (in USD) is

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

Taking logs gives

$$\ln(1 + R) = \ln(5.1/5.5) + \ln(1.9/1.6) = -0.076 + 0.172 = 0.096.$$

From (7.10) the mean and variance of the log return are

$$\mathrm{E}\, r_t = \mathrm{E}\, r_t^* + \mathrm{E}\, \Delta s_t \tag{7.11}$$

$$\mathrm{Var}(r_t) = \mathrm{Var}(r_t^*) + \mathrm{Var}(\Delta s_t) + 2\,\mathrm{Cov}(r_t^*, \Delta s_t). \tag{7.12}$$

See Tables 7.1–7.2 for an empirical illustration. Notice that a negative covariance (the foreign local return is high at the same time as the foreign currency depreciates) may reduce the variance of the return measured in domestic currency.

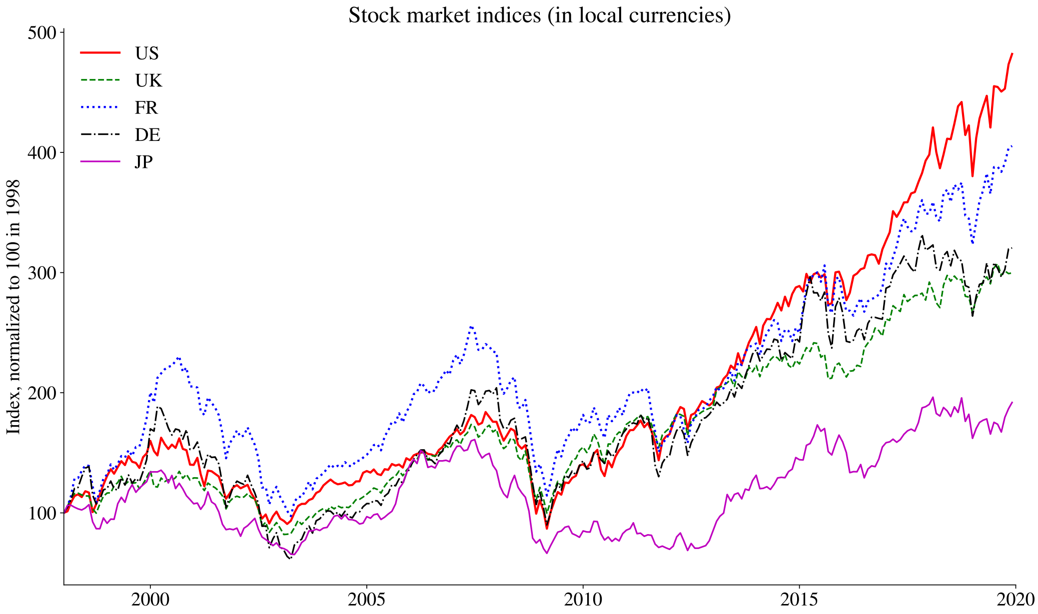


Figure 7.5: International stock market indices

	Local currency	Exchange rate	in USD
US	7.2	0.0	7.2
UK	5.0	−1.1	3.9
FR	6.4	0.1	6.5
DE	5.3	0.1	5.4
JP	3.0	0.8	3.8

Table 7.1: Contribution to the average (in %) log return for a US investor investing in different equity markets, 1998:01-2019:12

7.4 Hedging Exchange Rate Movements

International equity or bond investments often involve considerable exchange rate risk. It may be useful to hedge that risk. For instance, the investment strategy may be based on industry analysis ("pick promising pharma companies across the globe"), while the currency exposure is just unwanted risk (which requires a different type of analysis—and exchange rates are notoriously difficult to predict). Unless the covariance is very negative (as discussed above) this may motivate hedging the currency exposure.

The most common ways of hedging the exchange rate risk involve forward and option contracts (mostly for short horizons) or swap contracts (longer horizons). Alternatively, a

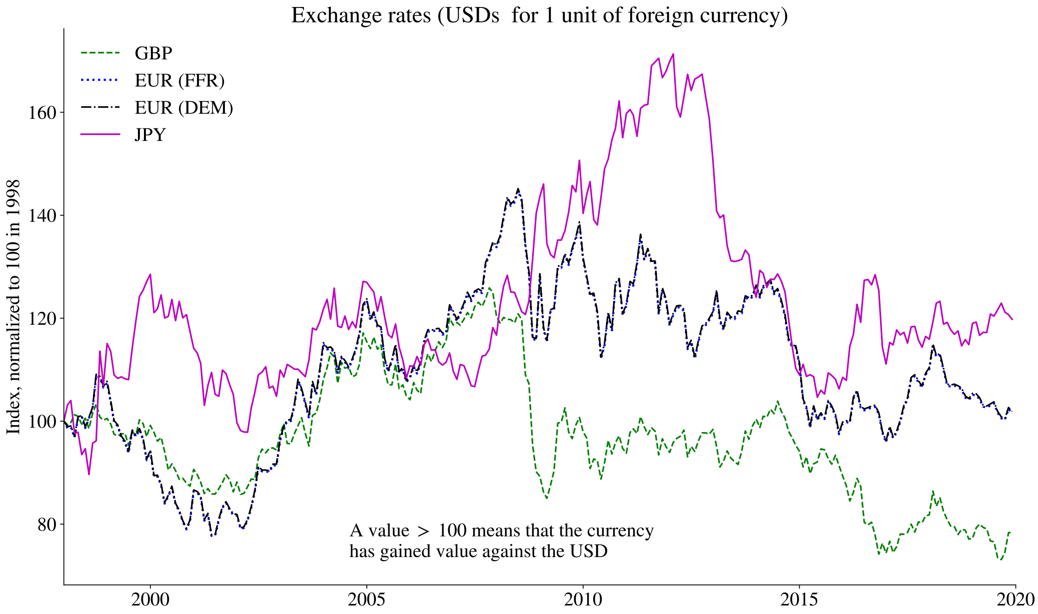


Figure 7.6: Exchange rate indices

	Local currency	Exchange rate	2*Cov	in USD
US	2.7	0.0	0.0	2.7
UK	2.0	0.8	0.1	2.9
FR	3.2	1.0	0.2	4.4
DE	4.5	1.0	0.2	5.7
JP	3.5	1.1	−1.6	3.0

Table 7.2: Contribution to the variance (in %) of the log returns for a US investor investing in different equity markets, 1998:01-2019:12

partial hedge is achieved by financing the investment by borrowing on the foreign market (in that way only the profit, not the entire investment, is exposed to exchange rate risk).

The perhaps most straightforward way to hedge the currency risk is by using a forward contract. For now, suppose we could lock in the period t exchange rate by entering a forward contract in $t-1$. If so, the return of the foreign investment (but measured in domestic currency) changes from (7.9) to

$$1 + R_t^{hedged} = (1 + R_t^*)\frac{F_{t-1}}{S_{t-1}}, \tag{7.13}$$

where the currency risk is eliminated.

The practical problem with (7.13) is that the foreign return, R_t^*, typically is not known

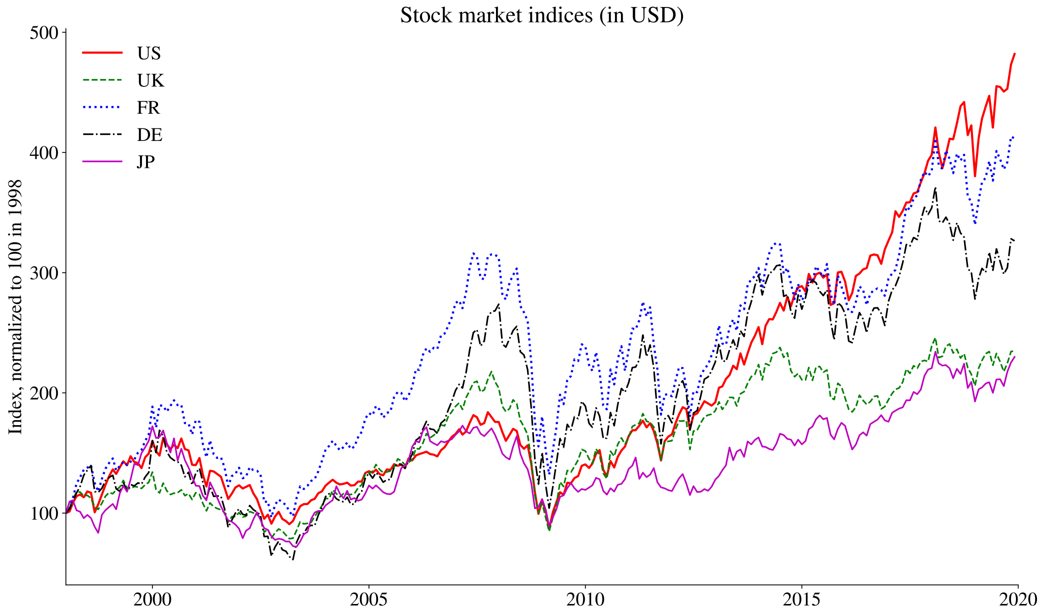


Figure 7.7: International stock market indices

in $t-1$, so we do not know how many units of currency to hedge via forward contracts. One possibility is to only hedge the investment, in which case the right hand side changes to $F_{t-1}/S_{t-1} + R_t^* S_t/S_{t-1}$.

Remark 7.17 *(*(7.13) when the foreign return is riskfree) Use the forward-spot parity (7.5) to substitute for the forward price in (7.13) to see that the hedged return then equals the domestic riskfree rate.*

7.5 Explaining Exchange Rates

Economic models of exchange rates can be thought of as trying to understand the "fundamental" value of currencies (similar to valuing a company according to the discounted sum of future dividends). This section briefly summarizes some of these ideas. It should be noticed, however, that most models of exchange rates only have explanatory power over longer horizons (5–10 years or longer).

7.5.1 Purchasing Power Parity and the Real Exchange Rate?

The basic idea is that a product should *cost the same at home and abroad* (when measured in a common currency). If this is not the case, then (goods) arbitrage will take place, driving down demand for the currency of the more expensive country which leads to an depreciation of its currency.

The strong assumption about goods arbitrage can be relaxed by instead assuming that goods may differ across countries, but that the import/export demand is somewhat price elastic. The *real exchange rate* (the relative price of foreign and domestic goods, measured in the domestic currency) is often used as an indicator of the competitiveness of a country. If the domestic price is too high, then export will decrease and import will increase, leading to a trade deficit. This puts pressure on the exchange rate in the same way as discussed above. The mechanism is thus that the real exchange rate puts pressure on the (nominal) exchange rate.

Empirical tests strongly refutes this set of theories for price and exchange rate *levels*, but may work reasonably well for changes over the long run (10+ years). In particular, it points at the important link between inflation (which drives up prices) and depreciations, which is a well established fact over longer runs. In the short run, the causality seems to be the reverse: (nominal) exchange rate movements cause movements in the real exchange rate (competitiveness).

An observation: price levels (measured in a common currency) are higher in rich countries. Once we adjust for that,we get a better measure of over/under valuation of the currency.

7.5.2 Interest Rates?

The exchange rate often appreciates when the central bank raises the interest rate. This typically happens very quickly. One possible explanation is financial flows: if international investors want to benefit from the higher interest rates, then they first need to buy the currency. However, if we were to believe in UIP then an investor needs to buy the currency before it has appreciated fully. Otherwise, the higher interest rate will be offset with a future depreciation. In short, the interest rate hike causes an immediate appreciation, followed by a slow depreciation.

Empirical tests suggests that high interest rate currencies can continue to appreciate for several years (this forms the basis for carry trades), but that they typically eventually suffer a sudden depreciation.

7.5.3 Transactions? (Business Cycles and Financial Flows)

The business cycle theory for exchange rates goes back to first principles to ask the question: why do we hold a currency (cash or cash-like assets)? After all, cash is typically not a good savings instrument (cash is eroded by inflation and there are typically better investment vehicles). Some cash is held because some people want to avoid banks (distrust of bank, fear of taxation and other legal issues), which seems to be an important driver of demand for large denomination bills. This may historically have had an effect on exchange rates, but less so today.

Instead, the key use of a currency is that it facilitates transactions, which suggests that both business cycle conditions (which drive the transaction volumes for goods and services) and financial flows are the most important factors behind exchange rates.

Empirical tests of these models find that also they have some explanatory power over longer horizons.

Chapter 8

Performance Analysis

Reference: Elton, Gruber, Brown, and Goetzmann (2014) 25 and 26
More advanced material is denoted by a star (*). It is not required reading.

8.1 Performance Evaluation

8.1.1 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

$$R^e_{it} = \alpha_i + b_i R^e_{mt} + \varepsilon_{it}, \text{ where} \qquad (8.1)$$
$$\mathrm{E}\,\varepsilon_{it} = 0 \text{ and } \mathrm{Cov}(R^e_{mt}, \varepsilon_{it}) = 0.$$

Example 8.1 *(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	0.00	1.00	0.00	10.00	18.00
p	1.00	0.90	14.00	10.00	21.41
q	5.00	1.30	3.00	18.00	23.59

Table 8.1: Basic facts about the market and two other portfolios, α, β, and Std(ε) are from CAPM regression: $R^e_{it} = \alpha + \beta R^e_{mt} + \varepsilon_{it}$

8.1.2 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

$$SR_p = \mu^e_p / \sigma_p. \qquad (8.2)$$

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 8.2 *(Performance measure) From Example 8.1 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

$$M^2_p = \mu^e_{p*} - \mu^e_m \text{ (or } \mu_{p*} - \mu_m), \qquad (8.3)$$

	SR	M^2	AR	Treynor	T^2
m	0.56	0.00		10.00	0.00
p	0.47	−1.59	0.07	11.11	1.11
q	0.76	3.73	1.67	13.85	3.85

Table 8.2: Performance Measures

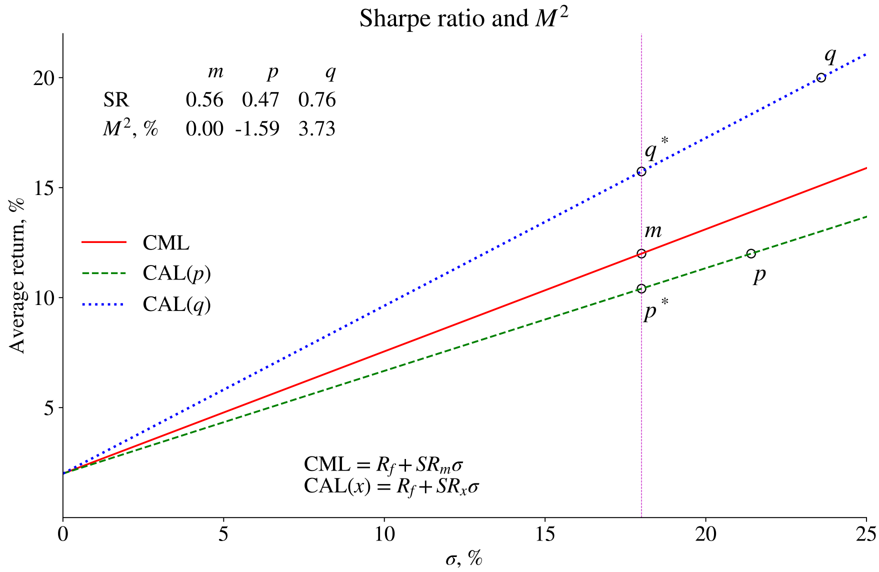


Figure 8.1: Sharpe ratio and M^2

where μ^e_{p*} 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.

$$R_{p*} = aR_p + (1-a)R_f, \text{ with } a = \sigma_m/\sigma_p. \tag{8.4}$$

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

$$\mu^e_{p*} = a\mu^e_p = \mu^e_p\sigma_m/\sigma_p \tag{8.5}$$

$$\sigma_{p*} = a\sigma_p = \sigma_m. \tag{8.6}$$

The latter shows that R_{p*} indeed has the same volatility as the market. See Example 8.2 and Figure 8.1 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 (8.2) to write

$$
\begin{aligned}
M_p^2 &= SR_{p^*}\sigma_{p^*} - SR_m\sigma_m \\
&= \left(SR_p - SR_m\right)\sigma_m .
\end{aligned} \tag{8.7}
$$

The second line uses the facts that R_{p^*} has the same Sharpe ratio as R_p (see (8.5)–(8.6)) and that R_{p^*} has the same volatility as the market. Clearly, the portfolio with the highest Sharpe ratio has the highest M^2.

8.1.3 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

$$
AR_p = \alpha_p / \operatorname{Std}(\varepsilon_{pt}), \tag{8.8}
$$

where α_p is the intercept and $\operatorname{Std}(\varepsilon_{pt})$ is the volatility of the residual of a CAPM regression (8.1). (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 return per unit of extra volatility (standard deviation). For instance, a ratio 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

$$
SR_c^2 = \left(\frac{\alpha_p}{\operatorname{Std}(\varepsilon_{pt})}\right)^2 + SR_m^2 . \tag{8.9}
$$

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 8.2 for an illustration.

If the alpha is negative, and we rule out short sales, then (8.9) 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*

$$
IR_p = \frac{\operatorname{E}(R_p - R_b)}{\operatorname{Std}(R_p - R_b)}, \tag{8.10}
$$

where R_b is some benchmark return. The information ratio is similar to the appraisal ratio—although a bit more general. The denominator in 8.10 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. Instead, when R_f is the benchmark, then the information ratio equals the Sharpe ratio.

Proof. From the CAPM regression (8.1) we have

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

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

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

■

8.1.4 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

$$TR_p = \mu^e_p/\beta_p. \tag{8.11}$$

A higher Treynor's ratio is better.

The TR measure can be rephrased in terms of expected returns—and could then be 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^*

$$R_{p^*} = aR_p + (1-a)R_f, \text{ with } a = 1/\beta_p. \tag{8.12}$$

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

$$\mu^e_{p*} = a\mu^e_p = \mu^e_p/\beta_p \tag{8.13}$$
$$\beta_{p*} = a\beta_p = 1, \tag{8.14}$$

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

$$T^2_p = \mu^e_{p*} - \mu^e_m = \mu^e_p/\beta_p - \mu^e_m, \tag{8.15}$$

so the ranking (of fund p and q, say) in terms of Treynor's ratio and the T^2 are the same. See Example 8.2 and Figure 8.2 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 8.3 *(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. ■

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 $\delta\mu^e_i$ and the incremental variance is $\delta^2\sigma^2_i + 2\delta\sigma_{id} \approx 2\delta\sigma_{id}$, since δ^2 is very small. (The variance of R

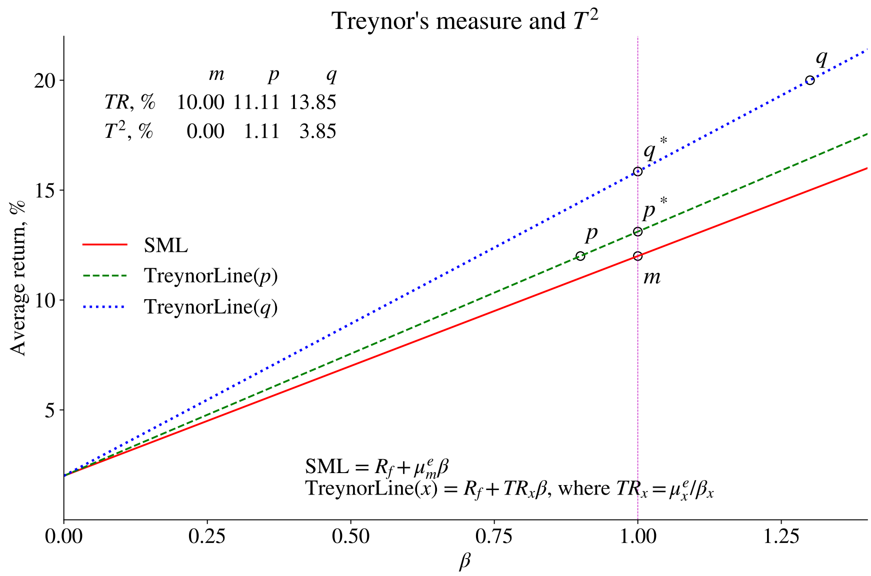


Figure 8.2: Treynor's ratio

is $\sigma_d^2 + \delta^2\sigma_i^2 + 2\delta\sigma_{id}$.) To a first-order approximation, the change (E $R_p - \text{Var}(R_p)k/2$) 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. ■

8.1.5 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

$$\begin{aligned} SR_p &= \frac{\alpha_p}{\sigma_p} + \text{Corr}(R_p, R_m)SR_m \\ AR_p &= \frac{\alpha_p}{\text{Std}(\varepsilon_{pt})} \\ TR_p &= \frac{\alpha_p}{\beta_p} + \mu_m^e. \end{aligned} \tag{8.16}$$

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 (8.16)*) Taking expectations of the CAPM regression (8.1) gives $\mu_p^e = \alpha_p + \beta_p \mu_m^e$, where $\beta_p = \mathrm{Cov}(R_p, R_m)/\sigma_m^2$. 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 (8.16) since

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

The AR_p in (8.16) 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	0.00	0.33	0.00		5.75	0.00
Putnam	0.21	0.33	−0.00	0.05	6.02	0.27
Vanguard	2.61	0.56	3.90	0.70	10.32	4.57

Table 8.3: Performance Measures of Putnam Asset Allocation: Growth A and Vanguard Wellington, weekly data 1999:01-2019:12 (annualized figures)

8.1.6 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) argue that market timing is similar to having a beta that is linear in the market excess return

$$\beta_i = b_i + c_i R^e_{mt}. \tag{8.17}$$

Using in a traditional market model (CAPM) regression, $R^e_{it} = a_i + \beta_i R^e_{mt} + \varepsilon_{it}$, gives

$$R^e_{it} = a_i + b_i R^e_{mt} + c_i (R^e_{mt})^2 + \varepsilon_{it}, \tag{8.18}$$

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), 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

$$\begin{aligned} R^e_{it} &= (a_i + \gamma_i z_{t-1}) + (b_i + \delta_i z_{t-1}) R^e_{mt} + \varepsilon_{it} \\ &= \theta_{i1} + \theta_{i2} z_{t-1} + \theta_{i3} R^e_{mt} + \theta_{i4} z_{t-1} R^e_{mt} + \varepsilon_{it}. \end{aligned} \tag{8.19}$$

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.

8.2 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 highlights how the performance has been achieved.

Grinblatt and Titman's measure (in period t) is

$$GT_t = \sum\nolimits_{i=1}^{n} (w_{i,t-1} - w_{i,t-2}) R_{it}, \tag{8.20}$$

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.

8.3 Performance Attribution

The performance of a fund depends on 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

$$R_p = \sum_{i=1}^{n} w_i R_{pi} \text{ and } R_b = \sum_{i=1}^{n} v_i R_{bi}, \tag{8.21}$$

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 $((\pm w_i R_{bi}))$ to get

$$R_p - R_b = \sum_{i=1}^{n} \left(w_i R_{pi} - v_i R_{bi}\right)$$
$$= \underbrace{\sum_{i=1}^{n} (w_i - v_i) R_{bi}}_{\text{allocation effect}} + \underbrace{\sum_{i=1}^{n} w_i \left(R_{pi} - R_{bi}\right)}_{\text{selection effect}}. \tag{8.22}$$

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 8.4 *(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 (8.22) 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).*

8.3.1 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.

8.4 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

$$R^e_{pt} = \sum_{j=1}^{K} b_j R^e_{jt} + \varepsilon_{pt}, \text{ with} \tag{8.23}$$
$$\sum_{j=1}^{K} b_j = 1 \text{ and } b_j \geq 0 \text{ for all } j.$$

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 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 8.3 for an example.

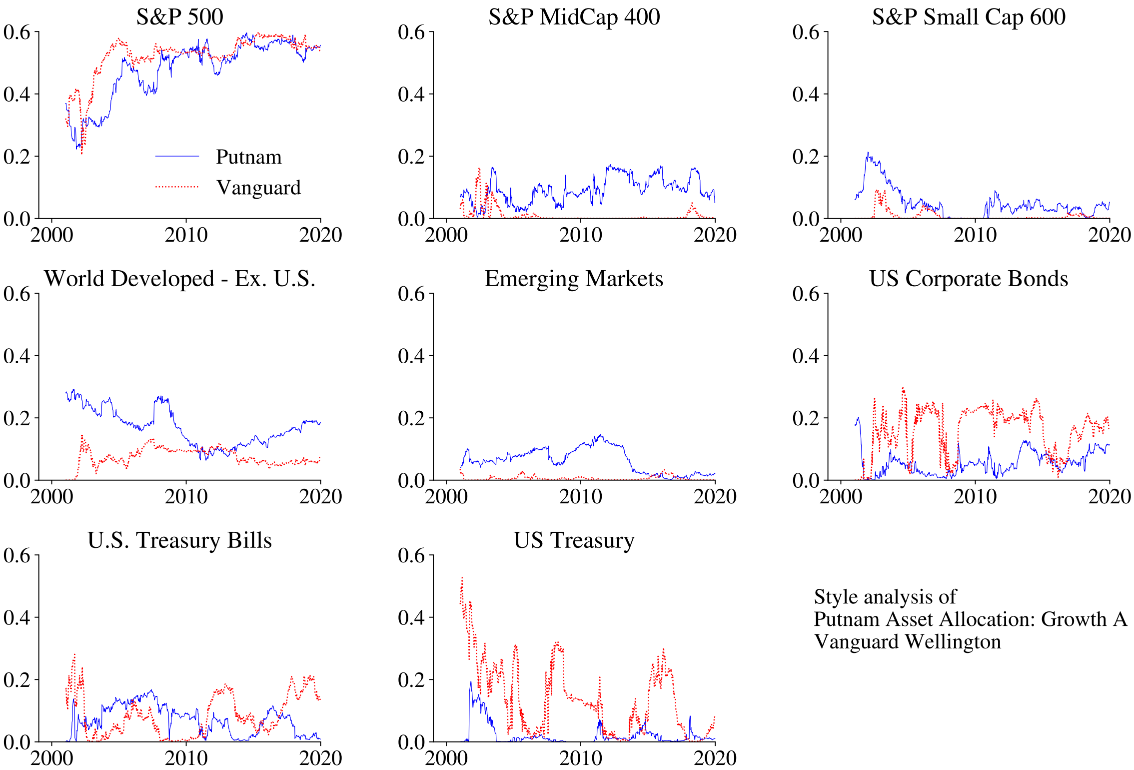


Figure 8.3: Example of style analysis, rolling data window

Chapter 9

Risk Measures

Reference: Hull (2009) 20; McDonald (2014) 31; McNeil, Frey, and Embrechts (2005); Alexander (2008)

9.1 Symmetric Dispersion Measures

9.1.1 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

$$\text{mean absolute deviation} = \mathrm{E}\,|R - \mu|. \tag{9.1}$$

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 9.1. 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 9.1 *(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* $N(0,1)$.

9.1.2 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 market 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

$$\mathrm{Var}(R_p - R_b) = (w - w_b)'\Sigma(w - w_b) \leq U, \tag{9.2}$$

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.

9.2 Downside Risk

9.2.1 Value at Risk

The mean-variance framework is often criticized for failing to distinguish between downside of the return distribution (considered to be risk) and upside (considered to be potential). The Value at Risk is one way of focusing on the downside.

Remark 9.2 *(Quantile of a distribution) The 0.05 quantile is the value such that there is only a 5% probability of a lower number,* $\mathrm{Pr}(R \leq quantile_{0.05}) = 0.05$.

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

$$\mathrm{Pr}(-R \geq \mathrm{VaR}_{95\%}) = 5\%. \tag{9.3}$$

Here, 95% is the confidence level of the VaR. For instance, if $\mathrm{VaR}_{95\%} = 18\%$, then we are 95% sure that we will not lose more than 18% of our investment. To convert the value

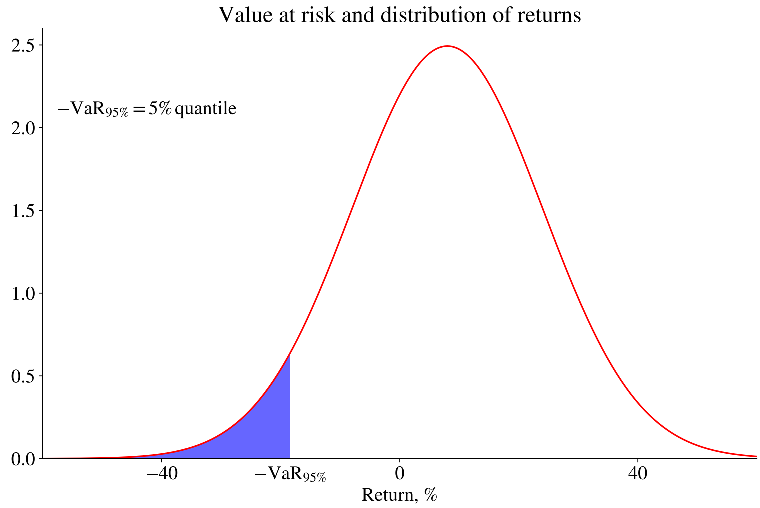


Figure 9.1: Value at risk

at risk into value terms (CHF, say), just multiply the VaR for returns with the value of the investment (portfolio).

Clearly, $-R \geq$VaR$_{95\%}$ is true when (and only when) $R \leq -$VaR$_{95\%}$, so (9.3) can be written

$$\Pr(R \leq -\text{VaR}_{95\%}) = 5\%. \tag{9.4}$$

This says that $-$VaR$_{95\%}$ is a number such that there is only a 5% chance that the return is below it. That is, $-$VaR$_{95\%}$ is the 0.05 quantile (5th percentile) of the return distribution. Using (9.4) allows us to work directly with the return distribution (not the loss distribution), which is often convenient. See Figure 9.1 for an illustration. If the return is normally distributed, $R \sim N(\mu, \sigma^2)$ then

$$\text{VaR}_{95\%} = -(\mu - 1.64\sigma). \tag{9.5}$$

Example 9.3 *(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 lose more than 18% of the investment over one day, that is, VaR$_{95\%}$ $= 0.18$.*

More generally, we can consider the confidence level α instead of just 0.95, so

$$\Pr(R \leq -\text{VaR}_\alpha) = 1 - \alpha, \text{ so} \tag{9.6}$$

$$\text{VaR}_\alpha = -[(1-\alpha)^{th} \text{ quantile of } R]. \tag{9.7}$$

In particular, if the return is normally distributed, $R \sim N(\mu, \sigma^2)$, then

$$\mathrm{VaR}_\alpha = -(\mu + c\sigma), \tag{9.8}$$

where c is the $(1-\alpha)^{th}$ quantile for a $N(0,1)$ distribution (for instance, $c = -1.64$ for $1-\alpha = 0.05$). See Figure 9.2 for an illustration.

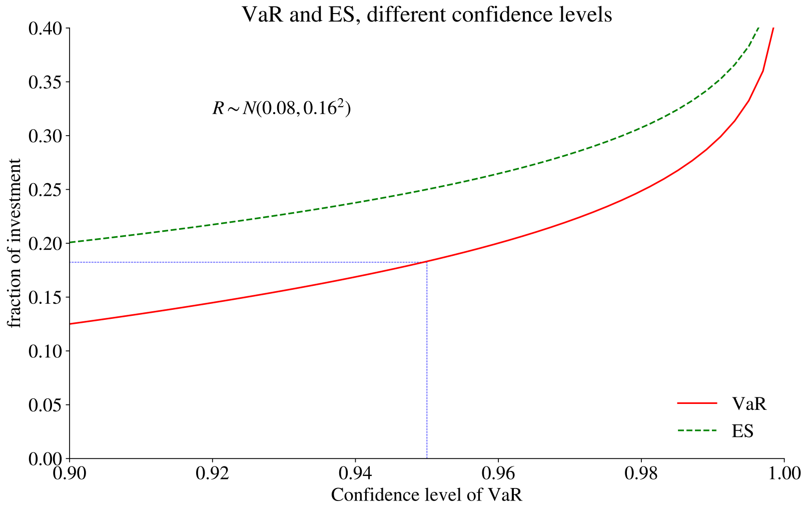


Figure 9.2: Value at risk, different probability levels

Remark 9.4 *(Critical values of $N(\mu, \sigma^2)$) If $R \sim N(\mu, \sigma^2)$, then the probability that $R \leq \mu + c\sigma$ is 5% for $c = -1.64$, 2.5% for $c = -1.96$, and 1% for $c = -2.33$.*

Example 9.5 *(VaR with $R \sim N(\mu, \sigma^2)$) If $R \sim N(\mu, \sigma^2)$ with $\mu = 8\%$ and $\sigma = 16\%$, then $VaR_{97.5\%} = -(0.08 - 1.96 \times 0.16) \approx 0.24$.*

Figure 9.3 shows the distribution and VaRs (for different probability levels) for the daily S&P 500 returns. Two different types of VaRs are shown: *(i)* based on a normal distribution and *(ii)* as the empirical VaR (from the empirical quantiles of the distribution).

Example 9.6 *(VaR and regulation of bank capital) Bank regulations have used 3 times the 99% VaR for 10-day returns as the required bank capital.*

Notice that the return distribution depends on the investment horizon, so a VaR 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

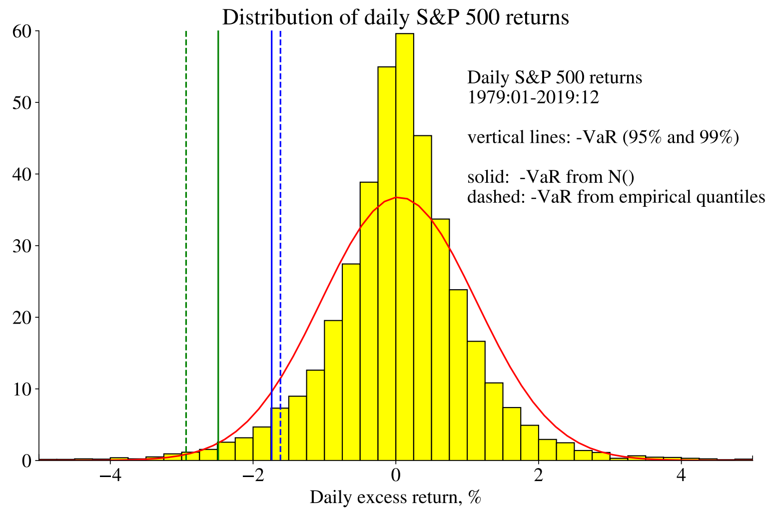


Figure 9.3: Return distribution and VaR for S&P 500

by making simplifying assumptions about the relation between returns in different periods (for instance, that they are iid).

Remark 9.7 *(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.*

Example 9.8 *(The London whale) The broad outline of the "London whale" (JPM) story is as follows: at the end of 2011, top management instructed the division to bring down the RWA (risk weighted asset) exposure to (various) credit derivatives. However, that would (a) have caused high execution costs and (b) the portfolio had recently performed well, so the division invented a new VaR method and pushed it through the Risk Office without the usual parallel testing. They went on to triple the positions (and lose $719 million in 2012Q1). Interestingly, the two VaR models show divergent paths for the value at risk.*

Example 9.9 *(Recommendations about risk budgeting). Here are some general recommendations about risk budgeting: (a) define risk limits and stick to them; (b) do not rely on a single risk measure; (c) be extra careful when changing risk measure; (d) do stress tests, eg. using the most extreme events during the last 15 years and/or identify which scenarios would hurt the most; (e) do not forget that risk changes (high volatility of individual assets, higher correlation among them).*

9.2.2 Backtesting a VaR model

While the results in Figure 9.3 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. The problem with that is illustrated in Figure 9.4.

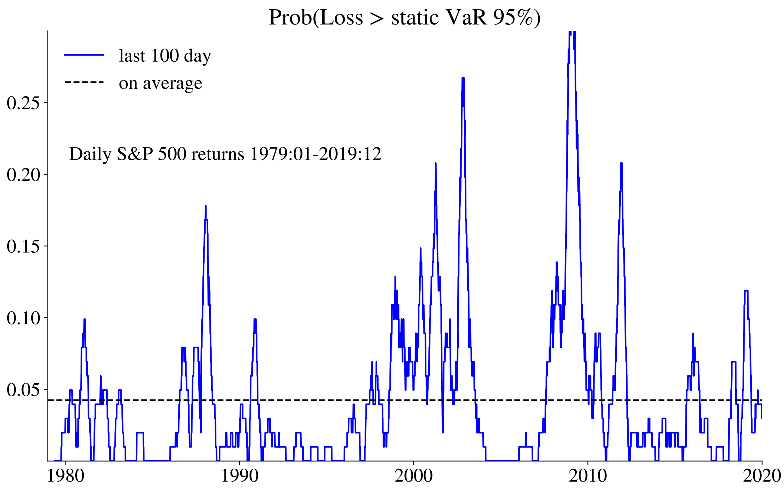


Figure 9.4: Backtesting a static VaR model on a moving data window

Figure 9.5 illustrates the VaR calculated from a time series 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. In short, the model is

$$\mu_t = \lambda \mu_{t-1} + (1-\lambda) R_{t-1} \tag{9.9}$$

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1-\lambda)(R_{t-1} - \mu_{t-1})^2. \tag{9.10}$$

The mean uses a "RiskMetrics" approach of updating yesterday's mean with yesterday's return. This is the same as a weighted average of past returns, but where recent data have higher weights that old data. The variance is a similar updating of yesterday's variance with the square of yesterday's surprise.

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 that is actually below (the negative of) this number. If the model is correct it should be 5%. We then repeat this for $\text{VaR}_{96\%}$: only 4% of the returns should be below (the negative of) this number.

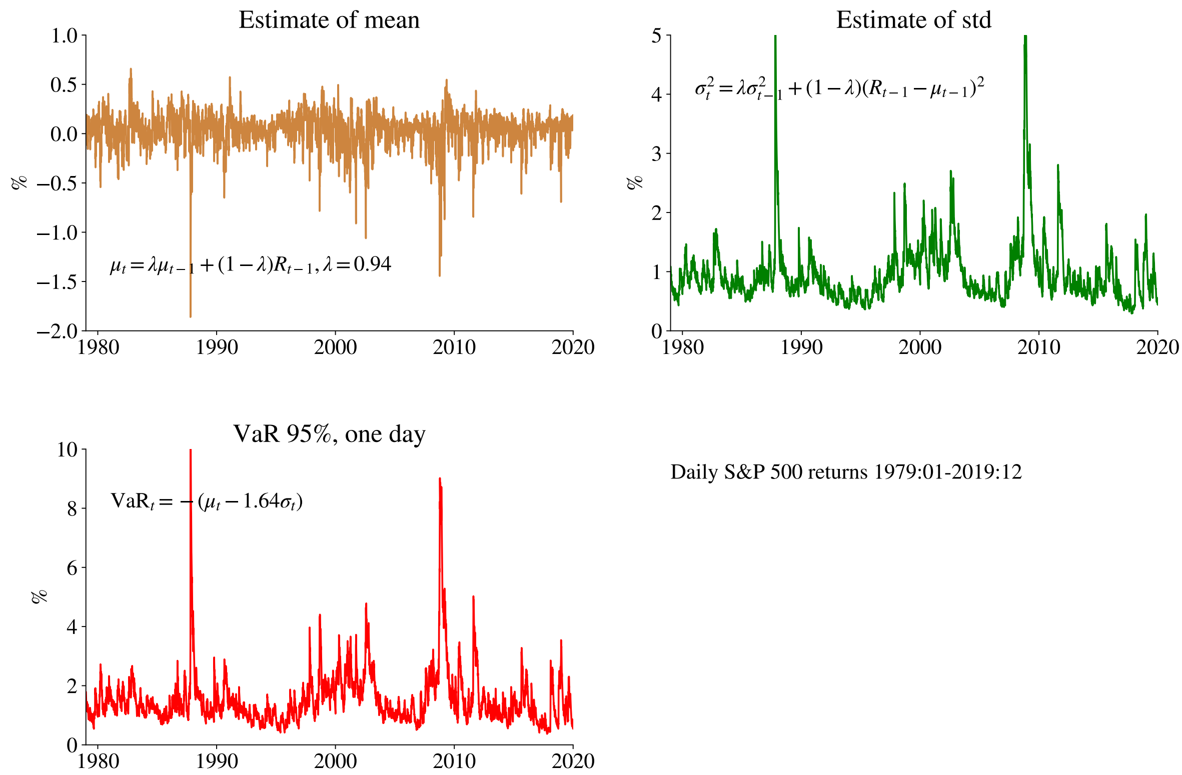


Figure 9.5: A dynamic VaR model

Figures 9.5–9.6 show results from backtesting a VaR model which assumes that one-day returns are normally distributed, but where the volatility is time varying. Clearly, this means that the VaR is also time varying: use (9.5) but allow σ (and less importantly, μ) to change from day to day. The evidence suggests that this model works relatively well at the 95% confidence level and that it is important to account for the time-varying volatility (or else there will be prolonged periods when the VaR performs poorly).

9.2.3 Value at Risk of a Portfolio

The general way of calculating the VaR of a portfolio is the same as for an individual asset (see above): first calculate (or estimate) the parameters of the distribution, then find the quantile.

However, in some special cases, there are ways to directly translate the VaR values of the individual assets to a portfolio VaR.

Remark 9.10 *Suppose the assets in the portfolio are jointly normally distributed with*

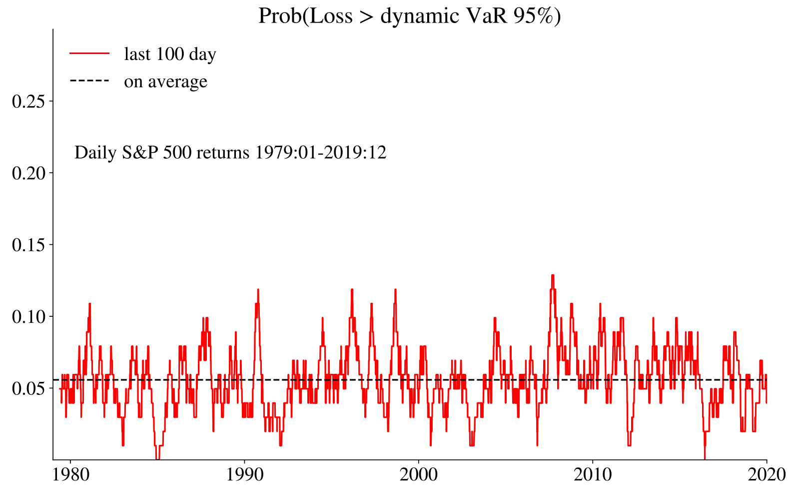


Figure 9.6: Backtesting a dynamic VaR model on a moving data window

zero means, so the VaR of asset i is $VaR_i = 1.64\sigma_i$. *Let* υ *be a vector where* $\upsilon_i = w_i VaR_i$, *where* w_i *is the portfolio weight. Then,* $VaR_p = [\upsilon' \operatorname{Corr}(R)\upsilon]^{1/2}$, *where* $\operatorname{Corr}(R)$ *is the correlation matrix of the assets. (To prove this, recall that* $VaR_p = 1.64\sigma_p$ *and that we can calculate* σ_p *from the* σ_i *values and correlations.)*

9.2.4 Index Models for Calculating the Value at Risk

Consider a multi-index model

$$
\begin{aligned}
R &= a + b_1 I_1 + b_2 I_2 + \ldots + b_k I_k + e, \text{ or} \\
&= a + b' I + e,
\end{aligned} \tag{9.11}
$$

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

$$
\begin{aligned}
\mu &= a + b' \mathrm{E}\, I \\
\sigma &= \sqrt{b' \operatorname{Cov}(I) b + \operatorname{Var}(e)},
\end{aligned} \tag{9.12}
$$

which can be used in (9.8), that is, an assumption of a normal return distribution. If the return is of a well diversified portfolio and the indices include the key market indices, then the idiosyncratic risk $\operatorname{Var}(e)$ 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 (9.11) 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 (9.11) 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 (9.8), is called the *delta-normal method*.

9.2.5 Expected Shortfall

While the value at risk is a useful risk measure, it has the strange property that it does not make a distinction between a loss that is just below the VaR level and a loss that is a lot below it. The VaR only cares about whether the outcome is in the tail of the return distribution, not how far out.

In addition, 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.) The *expected shortfall* has better aggregation properties.

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

$$\text{ES}_\alpha = -\text{E}(R|R \leq -\text{VaR}_\alpha). \tag{9.13}$$

See Figure 9.7 for an illustration.

See Table 9.1 for an empirical comparison of the VaR, ES and some alternative downside risk measures (discussed below).

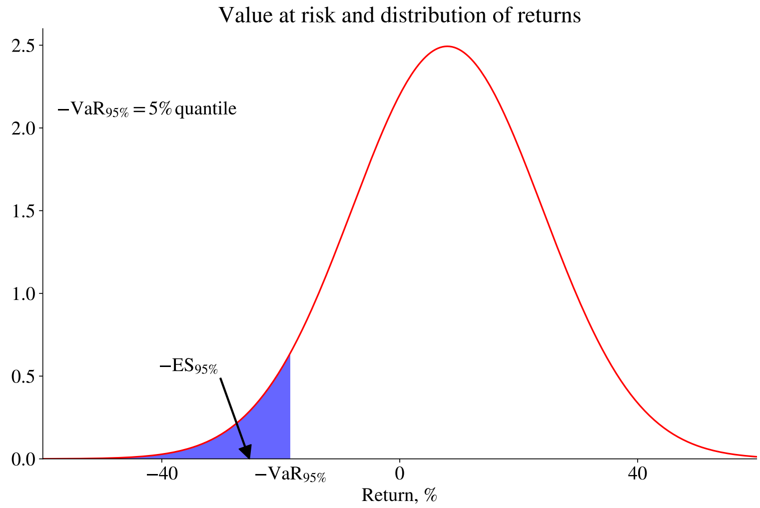


Figure 9.7: Value at risk and expected shortfall

	Small growth	Large value
Std	7.8	5.3
VaR (95%)	12.0	8.4
ES (95%)	16.8	11.5
SemiStd	5.4	3.5
Drawdown	79.4	59.3

Table 9.1: Risk measures of monthly returns of two stock indices (%), US data 1957:01-2019:12.

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

$$\text{ES}_\alpha = -\mu + \frac{\phi(c)}{1-\alpha}\sigma, \tag{9.14}$$

where $\phi()$ is the pdf of a $N(0,1)$ variable and c is the $1-\alpha$ quantile of a $N(0,1)$ distribution (for instance, -1.64 for $1-\alpha = 0.05$). See Figure 9.2.

Proof. (of (9.14)) If $x \sim N(\mu, \sigma^2)$, then it is well known that $\text{E}(x|x \leq b) = \mu - \sigma\phi(b_0)/\Phi(b_0)$ 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 $b = -\text{VaR}_\alpha = \mu + c\sigma$ so $b_0 = c$. Clearly, $\Phi(c) = 1-\alpha$, so $\text{E}(R|R \leq -\text{VaR}_\alpha) = \mu - \sigma\phi(c)/(1-\alpha)$. Multiply by -1. ■

Example 9.11 *(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 $ES_{97.5\%} = -0.08 + 0.16\phi(-1.96)/0.025 \approx 0.29$.*

Instead, to estimate the expected shortfall from the empirical return distribution, use

$$\text{ES}_\alpha = \frac{-1}{\sum_{t=1}^{T}\delta_t}\sum_{t=1}^{T}\delta_t R_t, \text{ where } \delta_t = \begin{cases} 1 \text{ if } R_t \leq -\text{VaR}_\alpha \\ 0 \text{ otherwise.} \end{cases} \tag{9.15}$$

This expression simply calculates the average $-R_t$ among those observations where $R_t \leq -\text{VaR}_\alpha$.

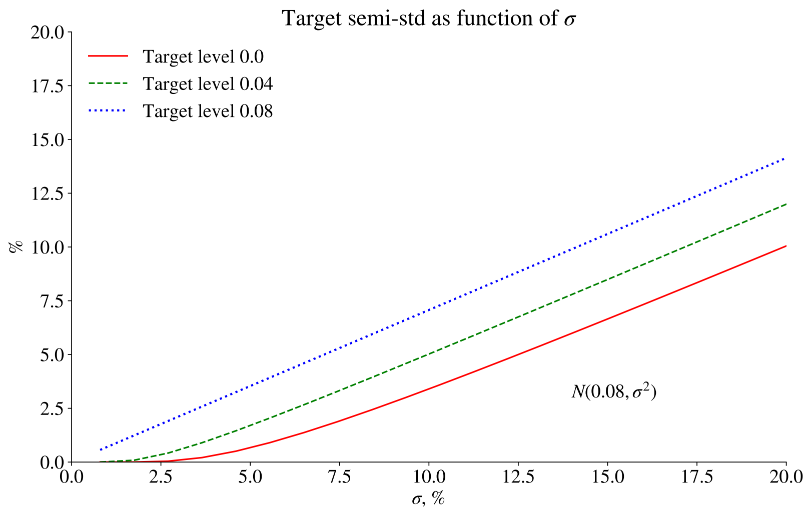


Figure 9.8: Target semivariance as a function of mean and standard deviation for a N(μ,σ^2) variable

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

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

The target semivariance (also called the lower partial 2nd moment) is defined as

$$\lambda(h) = \text{E}[\min(R - h, 0)^2], \tag{9.16}$$

where h is a "target level" chosen by the investor. Also, $\sqrt{\lambda(h)}$ with $h = \mu$ is called the semi-standard deviation.

In comparison with the variance

$$\sigma^2 = \text{E}(R - \text{E}\, R)^2, \tag{9.17}$$

the target semivariance differs in two aspects: *(i)* it uses the target level h as a reference point instead of the mean μ: and *(ii)* only negative deviations from the reference point are given any weight.

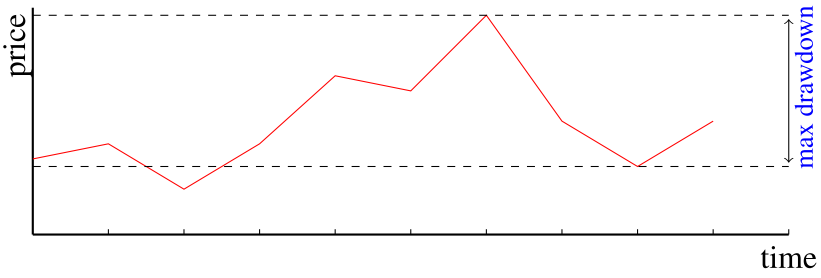


Figure 9.9: Max drawdown

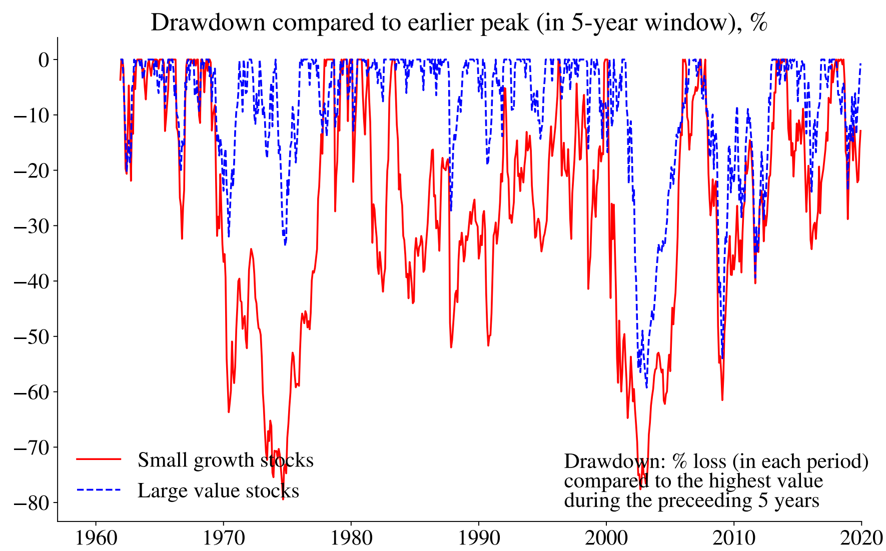


Figure 9.10: Drawdown

For a normally distributed variable, the target semivariance $\lambda_p(h)$ is increasing in the standard deviation (for a given mean)—see Remark 9.12. See also Figure 9.8 for an illustration.

Instead, to estimate the target semivariance from the empirical return distribution, use

$$\lambda(h) = \frac{1}{T}\sum\nolimits_{t=1}^{T} \delta_t (R_t - h)^2, \text{ where } \delta_t = \begin{cases} 1 \text{ if } R_t \leq h \\ 0 \text{ otherwise.} \end{cases} \quad (9.18)$$

This expression simply calculates the average of $\min(R_t - h, 0)^2$. (Warning: some analysts define $\lambda(h)$ by just including those observations when $R_t \leq h$. This means multiplying $\lambda(h)$ in (9.18) by $T/\Sigma_{t=1}^{T}\delta_t$. Conceptually, this is estimating $\mathrm{E}[(R-h)^2|R_t \leq h]$.)

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 9.9. 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 Figure 9.10 for an illustration.

Remark 9.12 *(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 9.8 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.

Remark 9.13 *(Sortino ratio) The Sortino ratio is an alternative to the Sharpe ratio (as a measure of performance). It is $(\mathrm{E}\,R-h)/\sqrt{\lambda(h)}$.*

See Table 9.2 for an empirical comparison of the different risk measures.

	Std	VaR (95%)	ES (95%)	SemiStd	Drawdown
Std	1.00	0.97	0.98	0.98	0.63
VaR (95%)	0.97	1.00	0.94	0.95	0.67
ES (95%)	0.98	0.94	1.00	0.98	0.66
SemiStd	0.98	0.95	0.98	1.00	0.66
Drawdown	0.63	0.67	0.66	0.66	1.00

Table 9.2: Correlation of rank of risk measures across the 25 FF portfolios (%), US data 1957:01-2019:12. The VaR and ES are based on the empirical return distribution. The max drawdown is calculated over a moving 5-year data window.

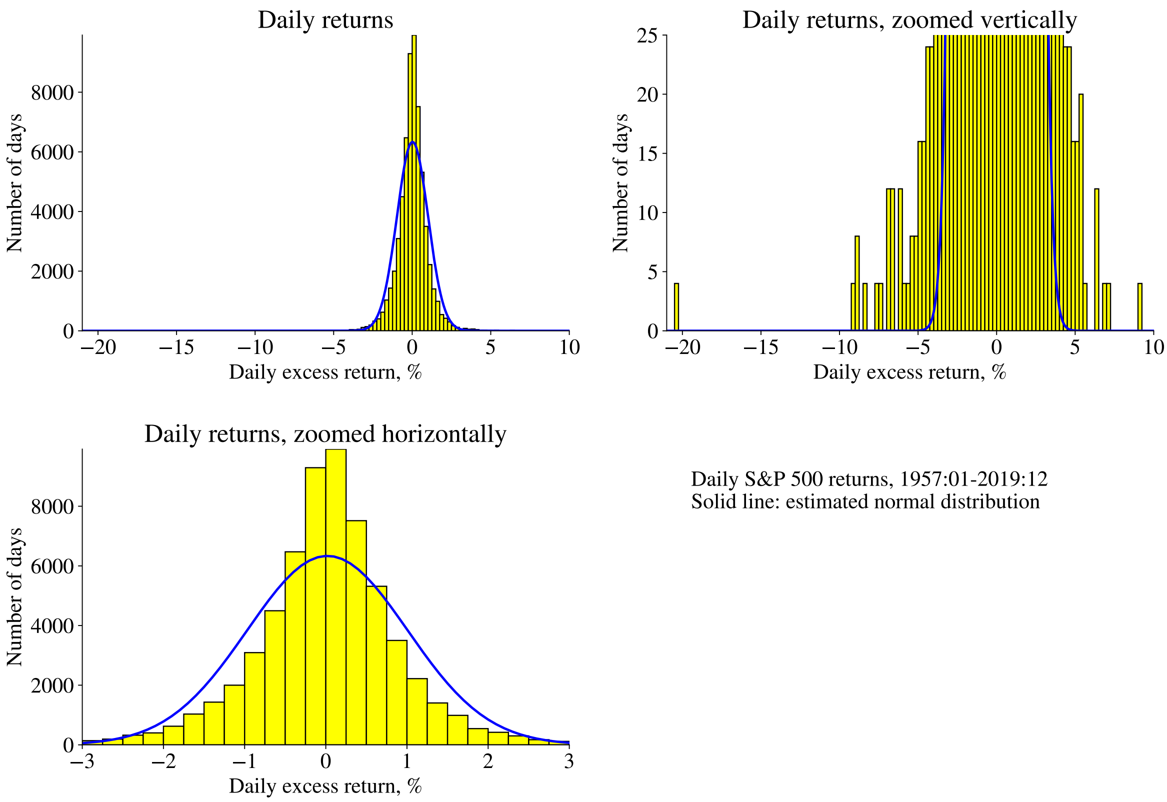


Figure 9.11: Distribution of daily S&P returns

9.3 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 9.11–9.13 for illustrations.

Remark 9.14 *(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 empirical as well as according to the theoretical distribution. For instance, if the 2th percentile*

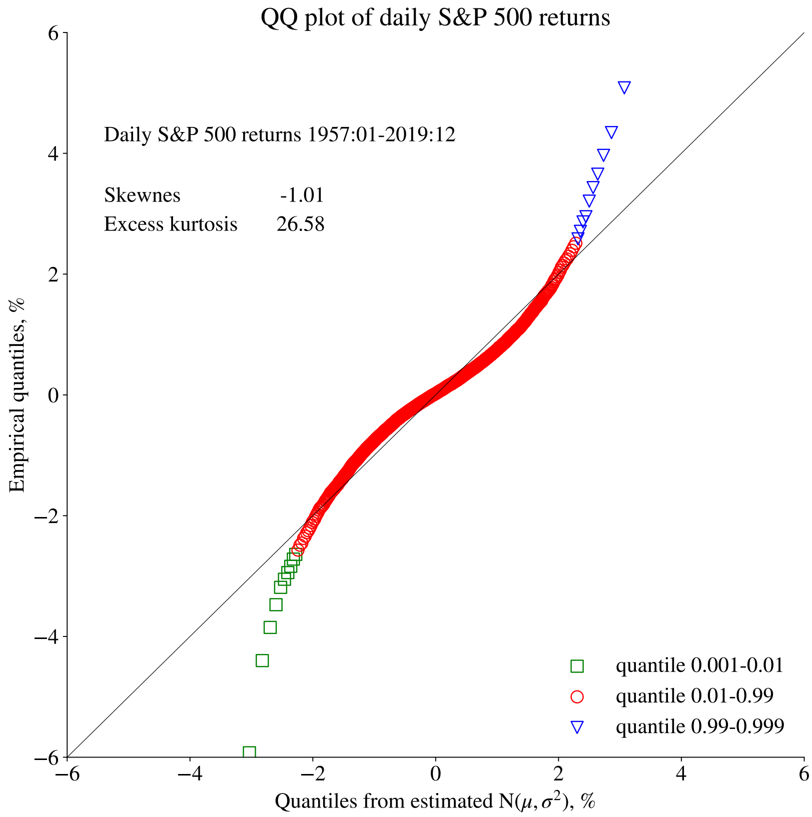


Figure 9.12: Quantiles of daily S&P returns

(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.

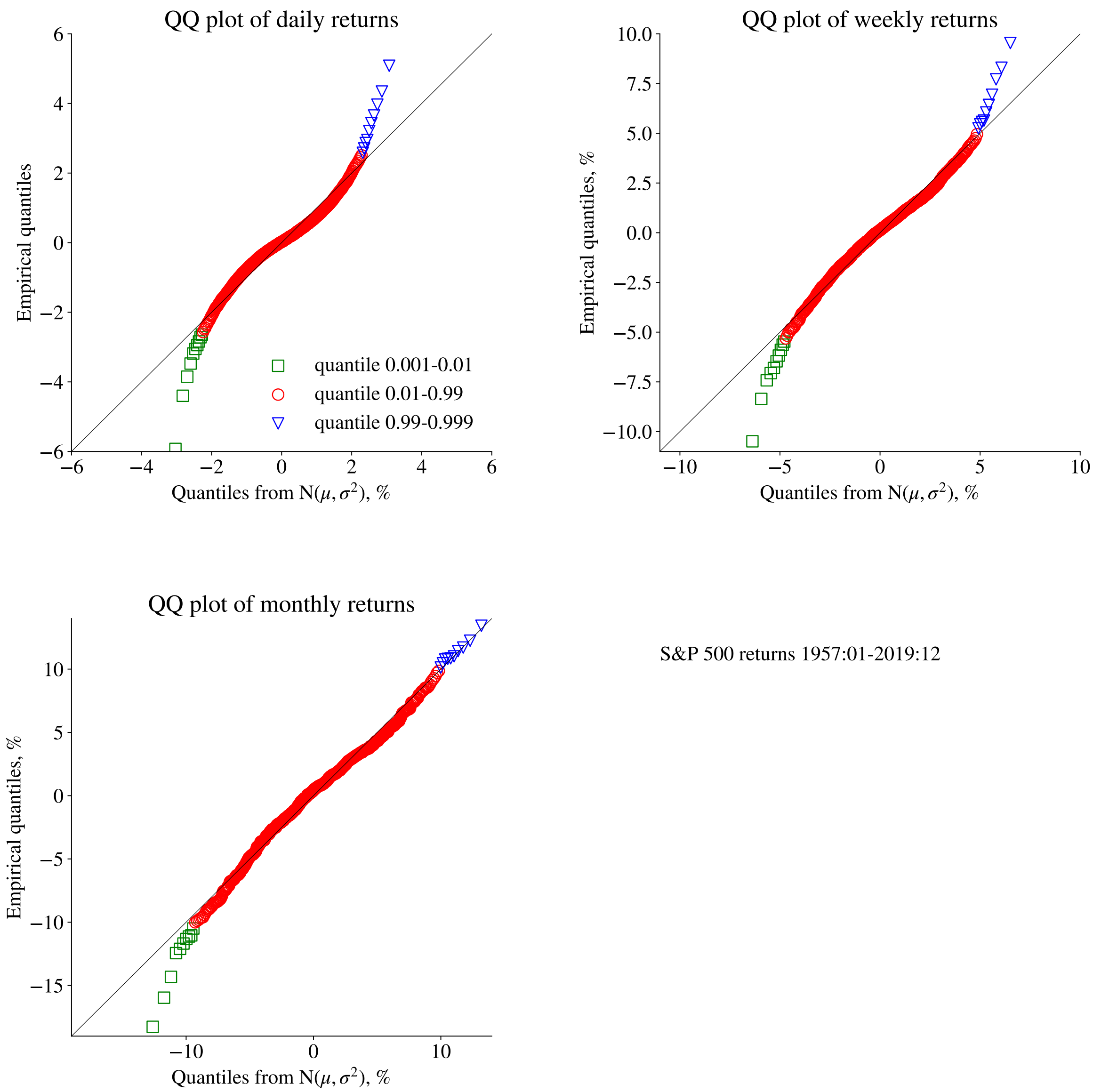


Figure 9.13: Distribution of S&P returns (different horizons)

Chapter 10

Utility-Based Portfolio Choice

Reference: Elton, Gruber, Brown, and Goetzmann (2014) 11 and 20

Additional references: Danthine and Donaldson (2005) 4–6; Huang and Litzenberger (1988) 4–5; Cochrane (2001) 9 (5); Ingersoll (1987) 3–5 (6)

Material with a star (*) is not required reading.

10.1 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 refer to basic economic utility theory.

10.1.1 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 $U(x)$.

In finance (and quite a bit of microeconomics that incorporates uncertainty), the key features of the utility functions that we use are as follows. Figure 10.1 gives an illustration.

First, utility is a function of a scalar argument, $U(x)$. This argument (x) can be end-of-period wealth, a consumption basket or the portfolio return. In one-period investment problems, this 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,

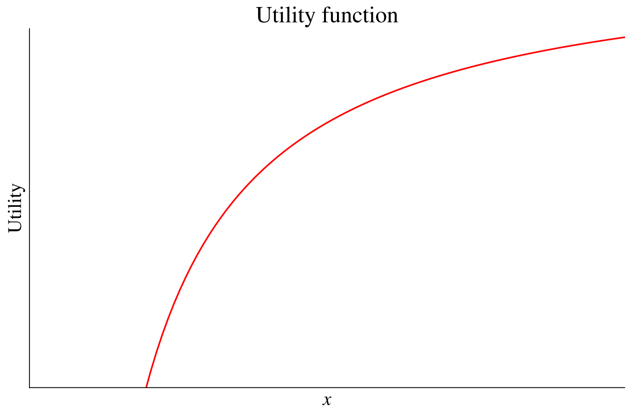


Figure 10.1: A utility function

$\mathrm{E}\, U(x)$. The reasons is that returns (and therefore wealth and consumption) are uncertain, so we need some way to rank portfolios at the time of investment (before the uncertainty has been resolved). For instance, if there are S possible outcomes ("states") $x_1, x_2, ..., x_S$ and outcome i has the probability π_i, then expected utility is

$$\mathrm{E}\, U(x) = \textstyle\sum_{i=1}^{S} \pi_i U(x_i). \tag{10.1}$$

Example 10.1 *($\mathrm{E}\, U(W)$) 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 $\mathrm{E}\, U(W) = 1/3 \times \ln 1 + 2/3 \times \ln 2 \approx 0.46$.*

Third, the functional form of the utility function (increasing, but concave) is such that more is better and uncertainty is bad (investors are risk averse).

10.1.2 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 (that is, we know what we like);
2. transitive: if A is better than B and B is better than C, then A is better than C (sounds like some 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).

10.1.3 Basic Properties of Utility Functions: (1) More is Better

The idea that *more is better* (non-satiation) is almost trivial. It means that the utility function is upward sloping. If $U(W)$ is differentiable, then this is the same as marginal utility being positive, $U'(W) > 0$.

Example 10.2 *(Logarithmic utility) $U(W) = \ln W$ so $U'(W) = 1/W$ (assuming $W > 0$).*

10.1.4 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.

In contrast, a linear utility function implies risk-neutrality, which we rule out. The reason is that investors seem to care about risk. If they did not, they would be equally happy holding a risky asset as a riskfree asset—as long as they have the same expected return. Conversely, they would also be happy to borrow as much as they could in order to hold an extremely leveraged portfolio of risky assets with high expected returns. This is not how typical investors behave. (Some may appear to do so, but they are often not gambling with their own money.)

Remark 10.3 *(CARA and CRRA utility functions) The CARA utility function is $U(W) = -e^{-kW}$ and the CRRA utility function is $U(W) = W^{1-k}/(1-k)$. (The abbreviations and names will be motivated later.)*

As an example, consider Figure 10.2. It shows a case where the portfolio (or wealth, or consumption,...) of an investor will be worth either x^- or x^+, each with a probability of 50%. The utility function shows risk aversion since the utility of getting the expected payoff for sure, $U(\mathrm{E}\,x)$, is higher than the expected utility from owning the uncertain asset

$$U(\mathrm{E}\,x) > 0.5U(x^{-}) + 0.5U(x^{+}) = \mathrm{E}\,U(x). \tag{10.2}$$

Remark 10.4 *(*Risk aversion and "marginal utility") Rearranging (10.2) gives*

$$U(\mathrm{E}\,x) - U(x^{-}) > U(x^{+}) - U(\mathrm{E}\,x),$$

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.

The (lowest) price (P) the investor is willing to sell this risky portfolio for is the certain amount of money which gives the same utility as $\mathrm{E}\,U\,(x)$, that is, the value of P that solves the equation

$$U(P) = \mathrm{E}\,U(x). \tag{10.3}$$

Notice that $U(P)$ is known, so not expectation is needed. This price P, called the *certainty equivalent* of the portfolio, is lower than the expected payoff

$$P < \mathrm{E}\,x = 0.5x^{-} + 0.5x^{+}. \tag{10.4}$$

(The result follows from $U(P) < U(\mathrm{E}\,x)$ and that $U()$ is an increasing function.) Again, see Figure 10.2 for an illustration.

Example 10.5 *(Certainty equivalent) Suppose you have a CRRA utility function and own an asset that gives either 0.85 or 1.15 with equal probabilities. What is the certainty equivalent (that is, the lowest price you would sell this asset for)? The answer is the P that solves*

$$\frac{P^{1-k}}{1-k} = 0.5\frac{0.85^{1-k}}{1-k} + 0.5\frac{1.15^{1-k}}{1-k}.$$

For instance, with $k = 0, 2, 5, 10,$ and 25 we have $P \approx 1, 0.977, 0.947, 0.912,$ and 0.875. Notice that P converges to 0.85 (the lowest outcome) as risk aversion increases.

This means that the *expected net return* on the risky portfolio that the investor demands is

$$\mathrm{E}\,R_x = \frac{\mathrm{E}\,x}{P} - 1 > 0, \tag{10.5}$$

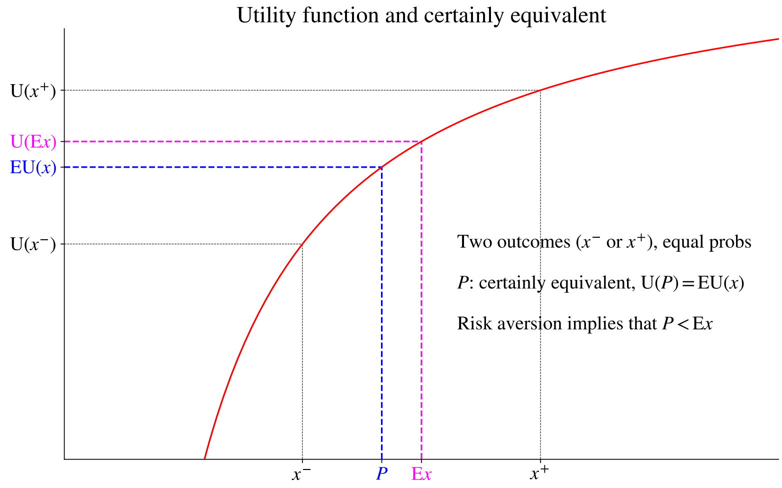


Figure 10.2: Certainty equivalent

which is greater than zero. This "required return" is higher if the investor is very risk averse (since P is lower). 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). Loosely speaking, we can think of $\mathrm{E}\, R_x$ as a *risk premium* (more generally, the risk premium is $\mathrm{E}\, R_x$ minus a riskfree rate). Notice that this analysis applies to the portfolio return (or wealth, or consumption,...), that is, the argument of the utility function—not to any individual asset. To 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 (not treated here).

Example 10.6 *(Risk premium in a simple case) Using the $k = 2$ case in Example 10.5 we get the expected net return (10.5)* $1/0.977 - 1 \approx 2.4\%$*, since* $\mathrm{E}\, x = 1$. *Instead, with* $k = 25$ *we get* $1/0.875 - 1 \approx 14.3\%$.

10.1.5 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 is related to the wealth level. (In contrast, when we use the portfolio return as the argument of the utility function, then this amounts to disregarding differences across wealth levels.)

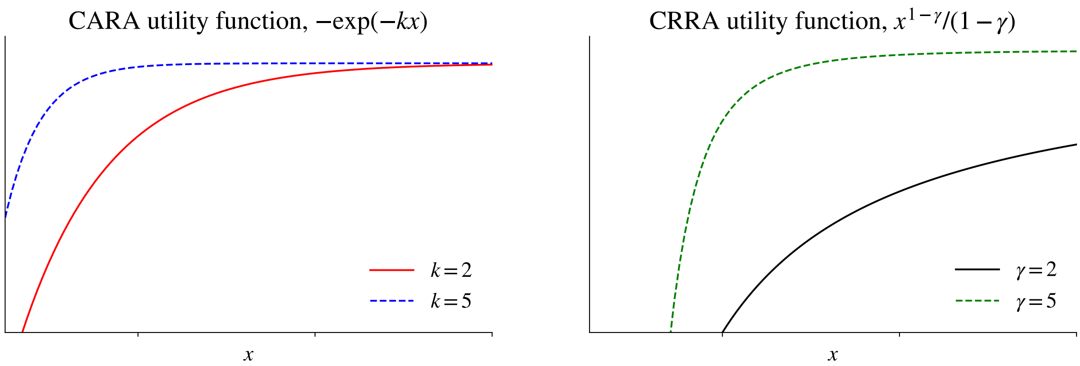


Figure 10.3: Examples of utility functions

First, define *absolute risk aversion* as

$$A(W) = \frac{-U''(W)}{U'(W)}, \tag{10.6}$$

where $U'(W)$ is the first derivative and $U''(W)$ the second derivative. Second, define *relative risk aversion* as

$$R(W) = WA(W) = \frac{-WU''(W)}{U'(W)}. \tag{10.7}$$

These two definitions are strongly related to the attitude towards taking risk (see below).

Figure 10.3 demonstrates two commonly used utility functions, and the following discussion outlines their main properties.

The *CARA utility function* (constant absolute risk aversion), $U(W) = -e^{-kW}$, is quite simple to use (in particular when returns are normally distributed), 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 would have a lower portfolio weight on risky assets.

Remark 10.7 *(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. Most evidence suggests that the CRRA utility function fits data reasonably well. For instance, historical data show no trends in portfolio weights or risk premia—in spite of investors having become much richer over time.

Remark 10.8 *(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.*

To understand the concepts of absolute and relative risk aversion, 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

$$\mathrm{E}\, U(W + Z) = U(W - P). \tag{10.8}$$

If Z is a small risk, then we can use a second order approximation and solve for the price as

$$P \approx A(W)\,\mathrm{Var}(Z)/2. \tag{10.9}$$

This says that the price the investor is willing to pay to avoid the risk Z is proportional to the *absolute risk aversion* $A(W)$.

Example 10.9 *(Willingness to pay to avoid a risk) Suppose the investor has a CARA utility function with* $A(W) = 5$ *and that* $\mathrm{Var}(Z) = 1$*. Then,* $P = 5 \times 1/2 = 2.5$.

Proof. (of (10.9)) 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 (10.9). ■

If we change the setting in (10.8)–(10.9) to make the risk proportional to wealth, that is $Z = Wz$ where z is the risk factor, then (10.9) directly gives

$$\begin{aligned}
P &\approx A(W)W^2\,\mathrm{Var}(z)/2, \text{ so} \\
P/W &\approx R(W)\,\mathrm{Var}(z)/2.
\end{aligned} \tag{10.10}$$

This says that the fraction of wealth (P/W) that the investor is willing to pay to avoid the risk (z) is proportional to the *relative risk aversion* $R(W)$.

Example 10.10 *(Willingness to pay to avoid a risk) Suppose the investor has a CRRA utility function with* $R(W) = 5$ *and that* $\text{Var}(z) = 0.2$. *Then,* $P/W = 5 \times 0.2/2 = 0.5$.

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.

10.2 Utility-Based Portfolio Choice and Mean-Variance Frontiers

10.2.1 Utility-Based Portfolio Choice with a Single Risky Asset

Suppose the investor maximizes expected utility from the portfolio return by choosing between a risky and a riskfree asset

$$\max_v \mathrm{E}\, U(R_p), \text{ with } R_p = vR_1^e + R_f. \tag{10.11}$$

The first order condition with respect to the weight on risky assets is

$$\begin{aligned} \frac{\partial \mathrm{E}\, U(vR_1^e + R_f)}{\partial v} &= 0 \text{ or} \\ \mathrm{E}[U'(vR_1^e + R_f) \times R_1^e] &= 0, \end{aligned} \tag{10.12}$$

where $U'(vR_1^e + R_f)$ is shorthand notation for the marginal utility, evaluated at $vR_1^e + R_f$. Notice that the order of E and ∂ are different in the first and second expressions. This is permissible since E defines a sum (and a derivative of a sum is the sum of derivatives, see below for a remark). Also, notice that the second expression is the expectation of the *product* of marginal utility and the excess return.

Remark 10.11 *(*Interchanging the order of* E *and* ∂*) Recall that a derivative of a sum equals the sum of derivatives. To illustrate this, assume 2 possible outcomes* (x^-, x^+) *with probabilities* π *and* $1-\pi$, *a choice variable* v *and some differentiable function* $f(v, x)$. *The expectation is* $\mathrm{E}\, f(v,x) = \pi f(v,x^-) + (1-\pi) f(v,x^+)$.*Differentiating gives*

$$\frac{\partial \mathrm{E}\, f(v,x)}{\partial v} = \pi \frac{\partial f(v,x^-)}{\partial v} + (1-\pi)\frac{\partial f(v,x^+)}{\partial v},$$

which is $\mathrm{E}\, \partial f(v,x)/\partial v$.

Clearly, the first order condition (10.12) defines one equation in one unknown (v), so it should 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). Explicit solutions are only possible in very simple cases.

Example 10.12 *(Portfolio choice with log utility and two states) Suppose* $U(R_p) = \ln(R_p + 1)$*, and that there is one risky asset and a riskfree 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* $\max_v \mathrm{E}\, U(R_p)$ *where*

$$\mathrm{E}\, U(R_p) = \pi \ln(vR^{e-} + R_f + 1) + (1 - \pi) \ln(vR^{e+} + R_f + 1).$$

The first order condition ($\partial \mathrm{E}\, U(R_p)/\partial v = 0$*) is*

$$\pi \frac{R^{e-}}{vR^{e-} + R_f + 1} + (1 - \pi) \frac{R^{e+}}{vR^{e+} + R_f + 1} = 0,$$

so we can solve for the portfolio weight as

$$v = -(1 + R_f) \frac{\pi R^{e-} + (1 - \pi) R^{e+}}{R^{e-} R^{e+}}.$$

For instance, with $(R_f, R^-, R^+) = (0, -10\%, 12\%)$ *and* $\pi = 0.5$*, we get* $v = 0.83$*. Instead, with* $R^+ = 10\%$*, then we get* $v = 0$*. (Yes, the log utility function implies low risk aversion so the weights are very sensitive to the average returns.) See Figure 10.4 for an illustration.*

Remark 10.13 *(*When to put all investments into the riskfree asset?) Suppose* $v = 0$ *would be an optimal decision, then the portfolio return equals the riskfree rate which is not random. The first order condition (10.12) can then be written*

$$\mathrm{E}[U'(R_f) R_1^e] = U'(R_f)\, \mathrm{E}\, R_1^e = 0$$

which holds zero only if $\mathrm{E}\, R_1^e = 0$*. 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?)*

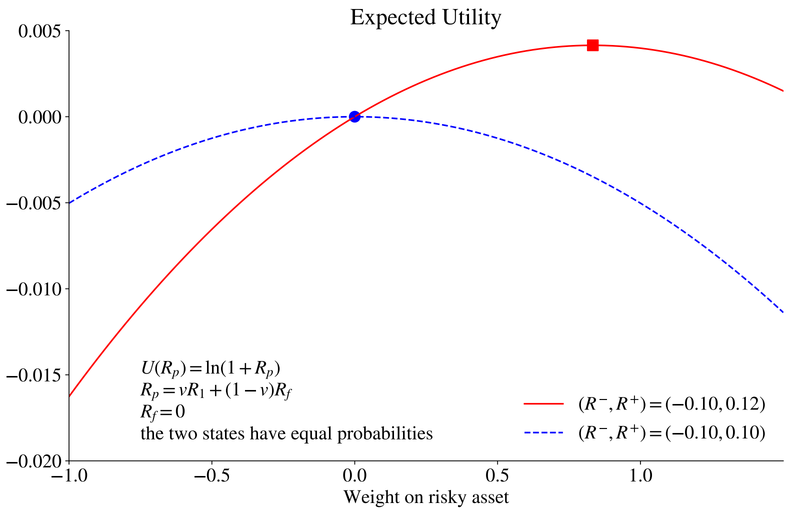


Figure 10.4: Example of portfolio choice with a log utility function

10.2.2 Utility-Based Portfolio Choice with Several Risky Assets

We now consider the case with n risky assets and a riskfree asset. The optimization problem is

$$\max_{v_1,v_2,\ldots} \mathrm{E}\, U(R_p), \text{ where} \tag{10.13}$$

$$R_p = \textstyle\sum_{i=1}^n v_i R_i^e + R_f. \tag{10.14}$$

where R_i^e is the excess return on asset i and R_f is a riskfree rate. Notice that the portfolio return can also be written $R_p = \Sigma_{i=1}^n v_i R_i + (1 - \Sigma_{i=1}^n v_i) R_f$, so the portfolio weights (on risky and riskfree) sum to one.

The first order conditions for the portfolio weights on the risky assets are

$$\frac{\partial\, \mathrm{E}\, U(R_p)}{\partial v_i} = 0 \text{ for } i = 1, 2, \ldots, n, \tag{10.15}$$

which defines n (possibly non-linear) equations in n unknowns: $v_1, v_2, \ldots, v_n$. Notice that calculating the expectation involves integrating over n dimensions. See Figure 10.5 for an illustration.

However, the (explicit or numerical) solution is often hard to obtain—so it would be convenient if we could simplify the problem.

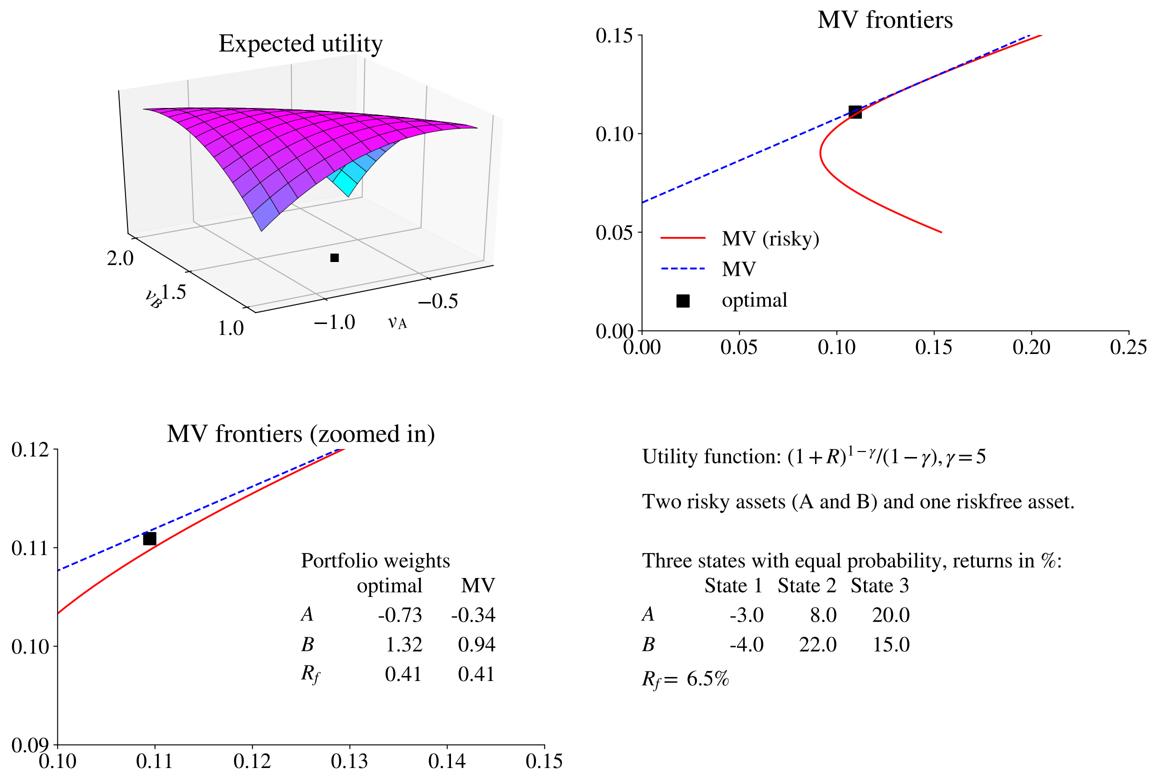


Figure 10.5: Example of when the optimal portfolio is (very slightly) off the MV frontier

10.2.3 Is the Optimal Portfolio on the Mean-Variance Frontier?

There are important cases where we can side-step most of the problems with solving the general portfolio choice problem (10.15)—since it can be shown that the portfolio will be on the mean-variance frontier (and we know how to calculate it).

The optimal portfolio must be on the mean-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, we solve

$$\max V(\mathrm{E}\, R_p, \mathrm{Var}(R_p)), \tag{10.16}$$
$$\text{with } \partial V()/\partial\, \mathrm{E}\, R_p > 0 \text{ and } \partial V()/\partial\, \mathrm{Var}(R_p) < 0.$$

In this case, we should interpret $V()$ as incorporating both the preferences and all relevant restrictions. As usual $\mathrm{E}\, R_p$ and $\mathrm{Var}(R_p)$ depend on the portfolio choice: $\mathrm{E}\, R_p = v'\mu$ and $v'\Sigma v$.

For an illustration, see Figure 10.6 which shows the iso-utility curves (curves with

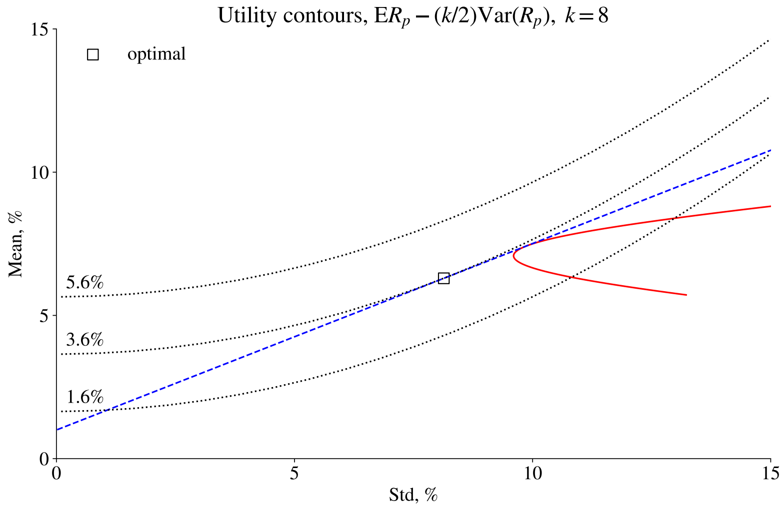


Figure 10.6: Iso-utility curves, mean-variance utility

equal utility) from a mean-variance utility function ($\mathrm{E}\,U(R_p) = \mathrm{E}\,R_p - (k/2)\,\mathrm{Var}(R_p)$). Whenever expected utility obeys (10.16) (not just for the mean-variance utility function) the iso-utility curves will look similar—so the optimum is on the mean-variance frontier. The intuition behind (10.16) is that an investor wants to move as far to the upper left as possible in Figure 10.6—but he/she is willing to trade off lower expected returns for lower volatility, that is, has iso-utility curves 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. Conditions for (10.16) to be true are discussed below.

See Figures 10.5–10.7 for examples of cases when (10.16) does not hold —and we do not get a mean-variance portfolio.

10.2.4 Special Cases

This section outlines special cases when the utility-based portfolio choice problem can be rewritten as in (10.16) (in terms of mean and variance only), so that the optimal portfolio is on the mean-variance frontier.

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.)

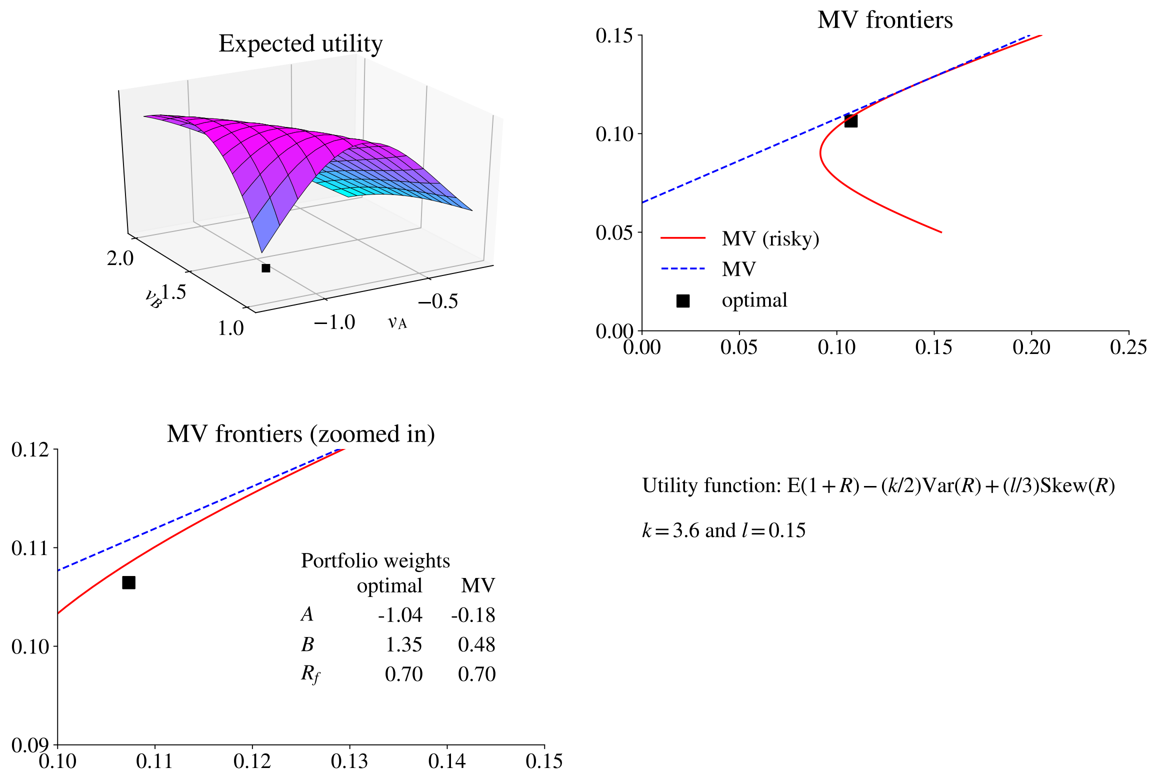


Figure 10.7: Example of when the optimal portfolio is (very slightly) off the MV frontier

Case 2: Quadratic Utility

If utility is quadratic in the return (or equivalently, in wealth)

$$U(R_p) = R_p - kR_p^2/2, \tag{10.17}$$

then expected utility can be written

$$\begin{aligned} \mathrm{E}\, U(R_p) &= \mathrm{E}\, R_p - k\, \mathrm{E}\, R_p^2/2 \\ &= \mathrm{E}\, R_p - k[\mathrm{Var}(R_p) + (\mathrm{E}\, R_p)^2]/2 \end{aligned} \tag{10.18}$$

since $\mathrm{Var}(R_p) = \mathrm{E}\, R_p^2 - (\mathrm{E}\, R_p)^2$. (We assume that all these moments are finite.) For $k > 0$ this function is decreasing in the variance, and increasing in the mean return as long as $k\, \mathrm{E}\, R_p < 1$. In this case, the optimal portfolio is on the mean-variance frontier.

The main drawback of this utility function is that we have to make sure that we are on the portion of the curve where expected utility is increasing in $\mathrm{E}\, R_p$ (below the so called "bliss point"). Moreover, the quadratic utility function has the strange property that the

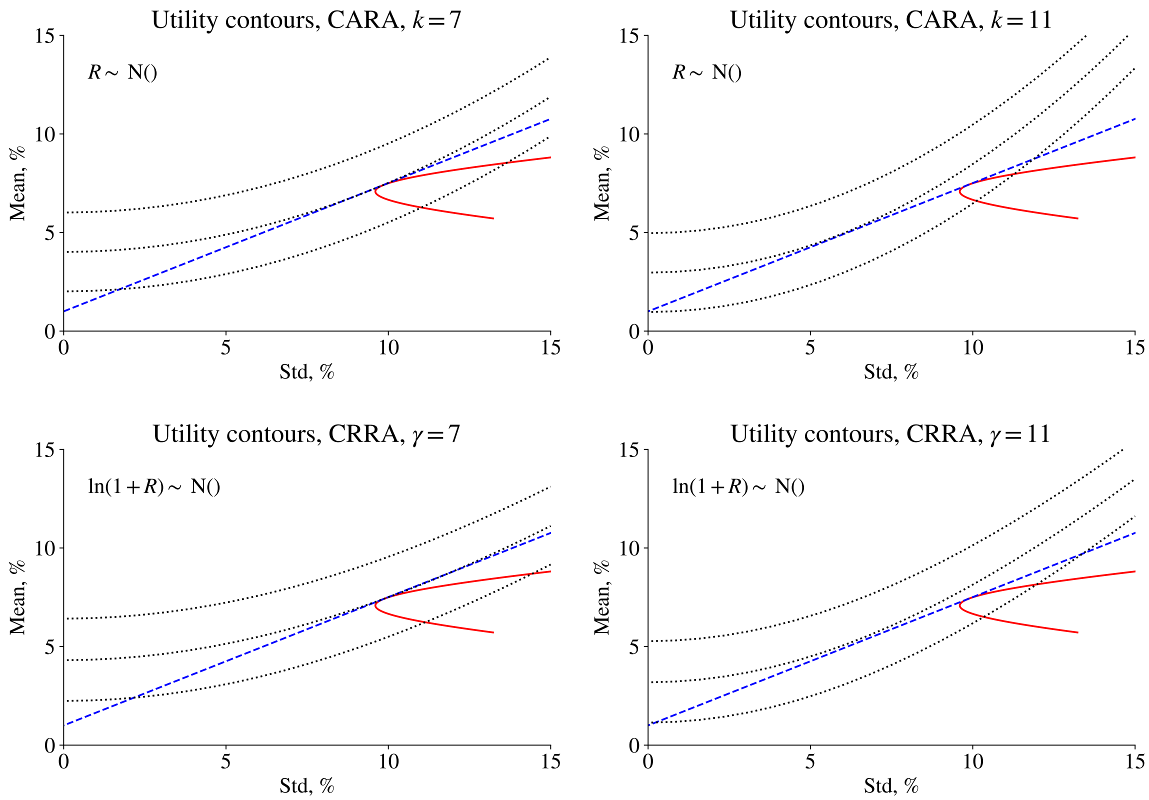


Figure 10.8: Contours with same utility level when returns are normally or lognormally distributed. The means and standard deviations (on the axes) are for the net returns (not log returns).

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 $\mathrm{E}\, U(R_p)$ 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 have a joint normal distribution). (An appendix provides details of the mathematics.)

The extra assumptions needed are that utility is strictly increasing ($U'() > 0$), displays risk aversion ($U''() < 0$), and utility must be defined for all possible outcomes. (The latter sounds trivial, but care is needed when specifying the utility function.)

Normally distributed returns should be considered as just an approximation for three reasons. First, limited liability means that the net return can never be below -100% (the asset price cannot be negative). However, such returns are possible in a normal distribution (although they may have very low probabilities). Second, 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. Thirds, 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.

As a special case of what happens when we combine a normal distribution with a valid utility function, consider the next proposition. Further examples/applications (for instance, using the Telser criterion) are discussed in a separate section below.

Proposition 10.14 *If returns are normally distributed, then maximizing the expected value a utility function with* constant absolute risk aversion $k > 0$ *(CARA)*

$$U(R_p) = -\exp(-R_p k)$$

is the same as solving a mean-variance problem. (The proof is in the appendix.)

See Figure 10.8 for an illustration of the case with CARA + normal distribution.

Case 4: CRRA Utility and Lognormally Distributed Portfolio Returns

Proposition 10.15 *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. (The proof is in the appendix.)*

See Figure 10.8 for an illustration of the case with CRRA + lognormal distribution.

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.)

10.3 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 the downside variation (considered to be risk) and the upside variation (considered to be potential). This section illustrates that normally distributed returns (typically) lead to minimum variance portfolios even if the portfolio selection model seems to be far from the standard mean-variance utility function.

10.3.1 Value at Risk

If the return is normally distributed, $R \sim N(\mu, \sigma^2)$, then the α Value at Risk (in terms of the return), VaR_α, is

$$\text{VaR}_\alpha = -(\mu + c\sigma), \tag{10.19}$$

where c is the $1-\alpha$ quantile of a N(0,1) distribution, for instance, -1.64 for 5% and -1.96 for 2.5%. If you instead want the Value at Risk in terms of value (not returns), multiply by the total value of the investment.

Example 10.16 *(VaR with $R \sim N(\mu, \sigma^2)$) If $\mu = 8\%$ and $\sigma = 16\%$, then $VaR_{95\%} =$* $-(0.08 - 1.64 \times 0.16) \approx 0.18$.

With normally distributed returns, the value at risk (10.19) is a strictly increasing function of the standard deviation (and the variance). In this case, the portfolio that minimizes the VaR at a given average return (minVaR st. $\mathrm{E}\, R_p = \mu^*$) will give a portfolio on the mean-variance frontier.

10.3.2 The Telser Criterion

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 (V^*)

$$\max_{v_i} \mathrm{E}\, R_p \text{ st. } \text{VaR}_\alpha < V^*. \tag{10.20}$$

When returns are normally distributed, (10.19) shows that the restriction can be written

$$\begin{aligned} -(\mathrm{E}\, R_p + c\sigma_p) &< V^*, \text{ so} \\ \mathrm{E}\, R_p &> -V^* - c\sigma_p. \end{aligned} \tag{10.21}$$

For instance, tilting the portfolio towards riskier assets is likely to increase $\mathrm{E}\, R_p$ but also σ_p, so there will be a trade-off (similar to a MV problem).

Example 10.17 *With a VaR confidence level of 95% and* $V^* = 0.1$*, then (10.21) gives* $\mathrm{E}\, R_p > -0.1 + 1.64\sigma_p$.

The optimization problem is (in general) solved by numerical methods, but the case of normally distributed returns can be handled analytically. This is illustrated in Figure 10.9. Any point above line defined by (10.21) 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 mean-variance frontier. A lower V^* limit is, of course, a tougher restriction.

The optimal portfolio is where the restriction intersects the mean-variance frontier: this gives the *highest possible* $\mathrm{E}\, R_p$ while obeying the restriction. This is clearly a point on the mean-variance frontier, which shows that the Telser criterion applied to normally distributed returns leads us to a mean-variance portfolio. If the restriction doesn't intersect, then there is no solution to the problem (the restriction is too demanding: the V^* is too low).

Since the solution is on the MVF, it is a mix of the tangency (market) portfolio denoted with subscript T (weight v) and the riskfree asset (weight $1 - v$). It is straightforward to show that the weight on the tangency portfolio that solves (10.20)

$$v = -\frac{R_f + V^*}{c\sigma_T + \mu_T^e}. \tag{10.22}$$

Example 10.18 *(Optimal portfolio. Telser) Let* $\mu_T^e = 6.5\%$*,* $\sigma_T = 10\%$ *and* $R_f = 1\%$*. The optimal portfolio with* $V^* = 10\%$ *is then*

$$v = -\frac{0.01 + 0.10}{-1.64 \times 0.1 + 0.065} \approx 1.11.$$

Instead, if the restriction is that VaR$< 4.5\%$*, then the weight is* $v \approx 0.55$.

Proof. (of (10.22)) The return on a portfolio which consists of the tangency portfolio and the riskfree is $R_p = vR_T^e + R_f$, so the mean and variance are $\mu_p = v\mu_T^e + R_f$ and

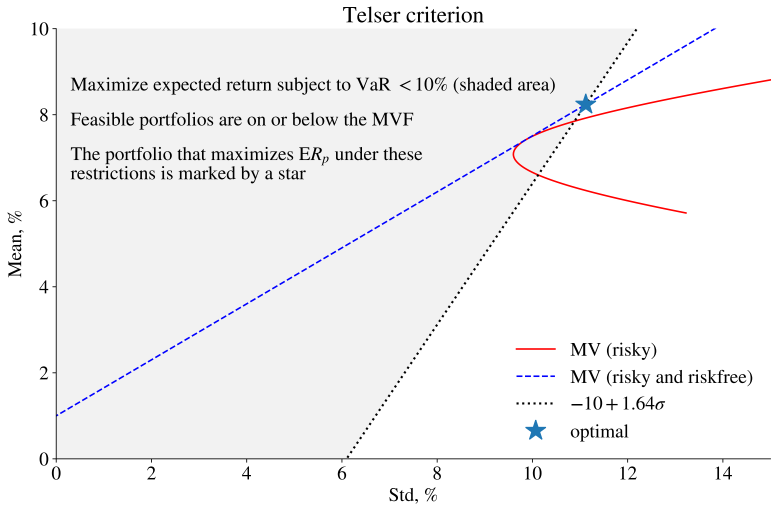


Figure 10.9: Telser criterion and VaR

$\sigma_p^2 = v^2\sigma_T^2$. Combining the mean return with the variance (to substitute for v, assuming $v \geq 0$) gives the CML

$$\mu_p = R_f + \frac{\mu_T^e}{\sigma_T}\sigma_p.$$

This equals the mean return required by the VaR restriction (10.21) when

$$\sigma_p = -\frac{R_f + V^*}{c + \mu_T^e/\sigma_T}.$$

Since $\sigma_p = v\sigma_T$ (assuming $v \geq 0$), the optimal portfolio weight on the tangency portfolio is (10.22). ■

10.3.3 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

$$\text{ES}_\alpha = -\mu + \frac{\phi(c)}{1-\alpha}\sigma, \tag{10.23}$$

where $\phi()$ is the pdf or a $N(0,1)$ variable and where c is the $1-\alpha$ quantile of a N(0,1) distribution, for instance, -1.64 for 5% and -1.96 for 2.5%.

Example 10.19 *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 (10.23) 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 a portfolio on the MV frontier

A "Telser type criterion" could, for instance, use the restriction $\mathrm{ES}_\alpha < b^*$, so

$$\mu_p > -b^* + \frac{\phi(c)}{1-\alpha}\sigma_p, \tag{10.24}$$

which define an area in a MV figure similar to that in Figure 10.9.

10.3.4 Target Semivariance

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

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

$$\lambda_p(h) = \mathrm{E}[\min(R_p - h, 0)^2], \tag{10.25}$$

where h is a "target level" chosen by the investor.

Suppose investor preferences are such that they like high expected returns and dislike the target semivariance (with a target level equal to the riskfree rate). This means that their expected utility can be written as

$$\begin{aligned} \mathrm{E}\, U(R_p) &= V(\mu_p, \lambda_p), \text{ with} \\ \partial(\mu_p, \lambda_p)/\partial\mu_p &> 0 \text{ and } \partial(\mu_p, \lambda_p)/\partial\lambda_p < 0. \end{aligned} \tag{10.26}$$

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 10.10 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. This is an intuitive result since, in this case, $\lambda_p(h)$ is an increasing function of the variance. See Figure 10.10 for a numerical illustration.

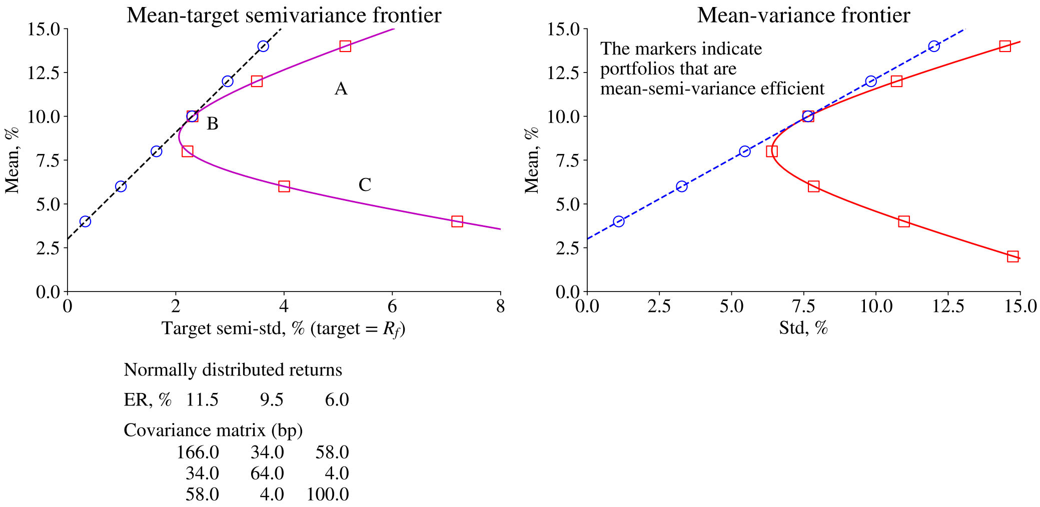


Figure 10.10: Target semivariance and expected returns

Remark 10.20 *(Target semivariance calculation for normally distributed variable*) For 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 $\lambda_p(\mu) = \sigma^2/2$, that is, when $h = \mu$.

10.4 Behavioural Finance

Reference: Elton, Gruber, Brown, and Goetzmann (2014) 20; 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.

10.4.1 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 key assumptions in utility-based portfolio choice might be reasonable. There are, however, 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 seem 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).

10.4.2 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 seem 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.

10.4.3 Prospect Theory

The *prospect theory* (developed by Kahneman and Tversky) tries 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.

10.5 Appendix: Extra Details on Portfolio Choice with Normally Distributed Returns

10.5.1 Case 3 and 4: Proofs*

Proof. (*of Proposition 10.14) 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

$$\mathrm{E}\, U(R_p) = \mathrm{E}\left[-\exp\left(-R_p k\right)\right] = -\exp\left[-\mathrm{E}\, R_p k + \mathrm{Var}(R_p) k^2/2\right].$$

Notice that the assumption of normally distributed returns is crucial for this result. Second, recall that if x maximizes $f(x)$, then it also maximizes $g[f(x)]$ if g is a strictly increasing function. The function $-\ln(-z)/k$ is defined for $z < 0$ and it is increasing in z. 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. ■

Proof. (*of Proposition 10.15) Notice that

$$\frac{\mathrm{E}(1+R_p)^{1-\gamma}}{1-\gamma} = \frac{\mathrm{E}\exp[(1-\gamma)r_p]}{1-\gamma}, \text{ where } r_p = \ln(1+R_p).$$

Since r_p is normally distributed, the expectation is (recall that if $x \sim N(\mu, \sigma^2)$, $\mathrm{E}\, e^x = e^{\mu+\sigma^2/2}$)

$$\frac{1}{1-\gamma}\mathrm{E}\exp[(1-\gamma)r_p] = \frac{1}{1-\gamma}\exp[(1-\gamma)\,\mathrm{E}\, r_p + (1-\gamma)^2\,\mathrm{Var}(r_p)/2].$$

Assume that $\gamma > 1$. The function $\ln[z(1-\gamma)]/(1-\gamma)$ is then defined for $z < 0$ and it is increasing in z. 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. ■

10.5.2 Case 3: Normally Distributed Returns, the Math of the General Case*

Remark 10.21 *(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

$$U(R_p) = \sum\nolimits_{n=0}^{\infty} \frac{1}{n!} \frac{d^n U(\mathrm{E}\,R_p)}{dW^n} \left(R_p - \mathrm{E}\,R_p\right)^n, \tag{10.27}$$

where $d^n U(\mathrm{E}\,R_p)/dW^n$ denotes the nth derivative of the utility function—evaluated at the point $\mathrm{E}\,R_p$. For instance, $d^2 U(\mathrm{E}\,R_p)/dW^2$ is the same as $U''(\mathrm{E}\,R_p)$.

Take expectations, but notice that $d^n U(\mathrm{E}\,R_p)/dW^n$ is not random, only the $(R_p - \mathrm{E}\,R_p)^n$ terms are. Also recall that $\mathrm{E}(R_p - \mathrm{E}\,R_p) = 0$ and that $\mathrm{E}(R_p - \mathrm{E}\,R_p)^2 = \mathrm{Var}(R_p)$. (As usual, $\mathrm{E}(R_p - \mathrm{E}\,R_p)^2$ should be understood as $\mathrm{E}[(R_p - \mathrm{E}\,R_p)^2]$.) Write out as

$$\mathrm{E}\,U(R_p) = U(\mathrm{E}\,R_p) + 0 + \frac{1}{2} U''(\mathrm{E}\,R_p)\,\mathrm{Var}(R_p) + \sum\nolimits_{n=3}^{\infty} \frac{1}{n!} \frac{d^n U(\mathrm{E}\,R_p)}{dW^n} \mathrm{E}\left(R_p - \mathrm{E}\,R_p\right)^n. \tag{10.28}$$

Remark 10.22 *(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 10.23 *(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 $n-1$, that is,* $1 \times 3 \times \ldots \times (n-3) \times (n-1)$.

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 (10.28) can be written

$$\mathrm{E}\,U(R_p) = U(\mathrm{E}\,R_p) + F(\mathrm{E}\,R_p)\,\mathrm{Var}(R_p), \tag{10.29}$$

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.

Chapter 11

CAPM Extensions and Multi-Factor Models

Reference: Elton, Gruber, Brown, and Goetzmann (2014) 14 and 16

11.1 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

$$\begin{aligned} R_i^e &= \beta_{im} R_m^e + \beta_{iF} R_F^e + \varepsilon_i, \text{ where} \\ \mathrm{E}\,\varepsilon_i &= 0, \mathrm{Cov}(R_m^e, \varepsilon_i) = 0, \mathrm{Cov}(R_F^e, \varepsilon_i) = 0. \end{aligned} \tag{11.1}$$

The pricing implication is a multi-beta model

$$\mu_i^e = \beta_{im}\mu_m^e + \beta_{iF}\mu_F^e. \tag{11.2}$$

Remark 11.1 *(When factors are not excess returns*) Equation 11.2 assumes that the factor can be expressed as an excess return—but that is not always the case. 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 (11.1) and (11.2). Second, we could instead reformulate the model by adding an intercept in (11.2) 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 (11.2), by using a cross-section of different assets ($i = 1, 2, \ldots$).*

We will consider several *theoretical* multi-factor models: the "CAPM with background risk" as well as a consumption-based model.

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). For instance, Fama and French (1993) estimate a three-factor model (capturing the market, the difference between small and large firms and the difference between value firms and growth firms) and show that it performs much better than CAPM. Also, 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.

11.2 CAPM with 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 perhaps also asset prices—at least as long as the background risk is correlated with some of the investable assets. The intuition is that the assets will be used to hedge against the background risk (and this demand will, in equilibrium, affect the prices).

11.2.1 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 a risky asset (with return R_i) which I will henceforth call "equity." He also has a background risk—in the form of an endowment (positive or negative) of an asset (with return R_H). In general, we would think of this as a non-traded asset. This could, for instance, be labour income or a house (positive endowment) or a combination of both. For a company, it could perhaps be 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

$$\mathrm{E}\,U(R_p) = \mathrm{E}\,R_p - \frac{k}{2}\,\mathrm{Var}(R_p), \text{ where} \tag{11.3}$$

$$R_p = vR_i + \phi R_H + (1 - v - \phi)R_f \tag{11.4}$$

$$= vR_i^e + \phi R_H^e + R_f. \tag{11.5}$$

Note that ϕ is the portfolio weight of the non-traded asset (which is not a choice variable—rather an "endowment") and $1 - \phi$ is the weight of the financial portfolio (riskfree plus equity). When the non-traded asset is a liability, then $\phi < 0$.

Use the budget constraint in the objective function to get

$$\mathrm{E}\,U(R_p) = v\mu_i^e + \phi\mu_H^e + R_f - \frac{k}{2}(v^2\sigma_{ii} + \phi^2\sigma_{HH} + 2v\phi\sigma_{iH}), \tag{11.6}$$

where σ_{ii} and σ_{HH} are the variances of equity and the non-traded asset respectively, and σ_{iH} is their covariance.

The first order condition for the weight on equity (the risky asset with return R_i), v, is $\partial\,\mathrm{E}\,U(R_p)/\partial v = 0$, that is,

$$0 = \mu_i^e - k(v\sigma_{ii} + \phi\sigma_{iH}), \text{ so}$$

$$v = \frac{\mu_i^e/k - \phi\sigma_{iH}}{\sigma_{ii}}. \tag{11.7}$$

The second term in the optimal portfolio weight, $-\phi\sigma_{iH}/\sigma_{ii}$ (also called the "hedging term") depends on how important the non-traded asset is in the portfolio (ϕ) and the covariance term (σ_{iH}). Clearly, if there is no non-traded asset ($\phi = 0$), then we are back in a traditional MV case.

Remark 11.2 *(Interpreting the hedging terms*) We can write the hedging term as $-\phi\beta$, where β is from a regression of the non-traded asset on the investable risky asset*

$$R_H^e = \alpha + \beta R_i^e + \varepsilon, \textit{ with } \beta = \sigma_{iH}/\sigma_{ii}.$$

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 non-traded asset, so a short position brings down the volatility of the total portfolio (including the endowment).

We can rewrite the portfolio return (11.4) in terms of the return on the *financial sub-*

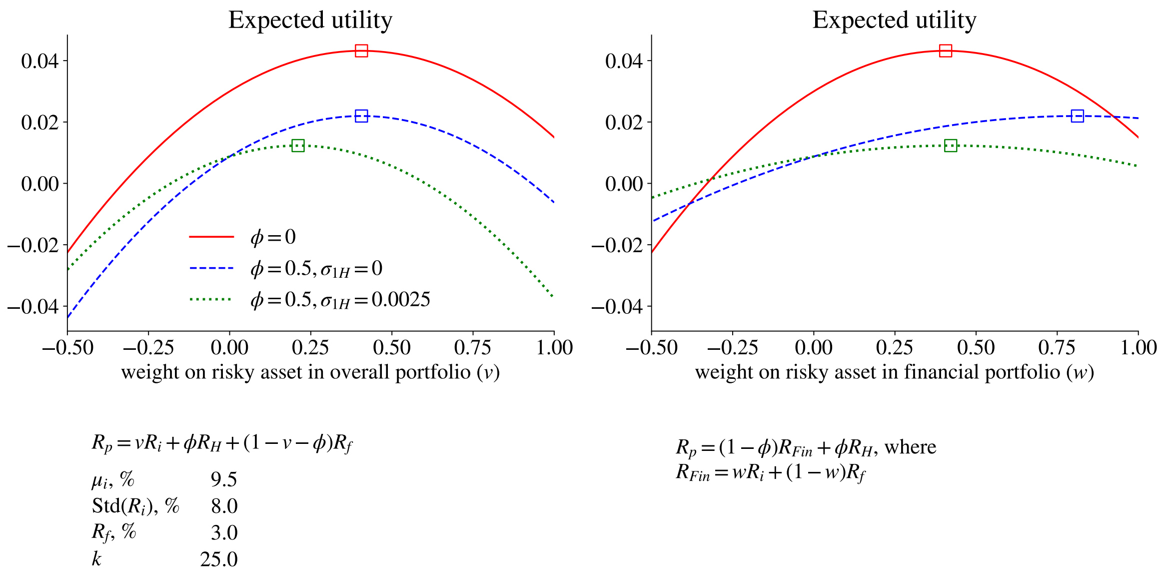


Figure 11.1: Portfolio choice with background risk

portfolio (R_{Fin}) and the non-traded asset as

$$R_p = (1 - \phi)R_{Fin} + \phi R_H, \text{ where} \tag{11.8}$$

$$R_{Fin} = wR_i + (1 - w)R_f. \tag{11.9}$$

Clearly, $(1 - \phi)w = v$, so (11.7) gives

$$w = \frac{\mu_i^e / k - \phi \sigma_{iH}}{(1 - \phi)\sigma_{ii}}. \tag{11.10}$$

Example 11.3 *(Portfolio choice with background risk) Figure 11.1 illustrates the v and w weights in several cases. Compared with the benchmark case A where there is no background risk ($\phi = 0$) we notice that*

	$\underline{v}$	$\underline{w}$
Case B ($\phi > 0.5, \sigma_{iH} = 0$)	*same as A*	*higher than A*
Case C ($\phi > 0.5, \sigma_{iH} > 0$)	*lower than B*	*lower than B*

Comparing cases A and B, we see that adding a non-traded asset which is uncorrelated with equity tilts the financial portfolio towards equity. Comparing cases B and C, we see that this effect is (at least partially) reversed if the non-traded asset is positively correlated with equity.

Example 11.4 *(Portfolio choice with a liability) Continuing Example 11.3, suppose now that the background risk is a liability (short position). Then*

	$\underline{v}$	$\underline{w}$
Case D ($\phi < 0, \sigma_{iH} = 0$)	*same as A*	*lower than A*
Case E ($\phi < 0, \sigma_{iH} > 0$)	*higher than D*	*higher than D*

Comparing cases A and D, we see that adding a liability risk that is uncorrelated with equity tilts the financial portfolio towards bonds. Comparing cases D and E, we see that a liability risk that is positively correlated with equity tilts the financial portfolio towards equity.

Several things can be noticed. First, *when the covariance is zero* ($\sigma_{iH} = 0$), 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 be tilted away from actual bonds (the riskfree asset).

Second, *when the covariance is positive* ($\sigma_{iH} > 0$) and we have a positive exposure to the background risk ($\phi > 0$), then the hedging term (second term) will tilt the financial portfolio away from equity and towards the riskfree asset. The intuition is that the overall portfolio now includes a lot of "equity like" assets, so the financial portfolio should be tilted towards the riskfree asset. 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 ($\sigma_{iH} < 0$, assuming a positive exposure, $\phi > 0$).

Example 11.5 *(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.*

Remark 11.6 *(Optimising over w directly*) Rewrite the portfolio return (11.4) as*

$$
\begin{aligned}
R_p &= w(1-\phi)R_i + (1-w)(1-\phi)R_f + \phi R_H \\
&= w(1-\phi)R_i^e + Z_f, \text{ where } Z_f = (1-\phi)R_f + \phi R_H.
\end{aligned}
$$

Optimizing the the objective function (11.3) wrt. w gives the same result as in (11.10).

11.2.2 Portfolio Choice with Background Risk: Several Risky Assets

	$\mu, \%$	Σ, bp		
		A	B	C
A	11.5	166	34	58
B	9.5	34	64	4
C	6.0	58	4	100

Table 11.1: Characteristics of the assets in the MV examples. Notice that $\mu, \%$ is the expected return in % (that is, $\times 100$) and Σ, bp is the covariance matrix in basis points (that is, $\times 100^2$).

With several risky assets the portfolio return is

$$R_p = v'R + \phi R_H + (1 - \mathbf{1}'v - \phi)R_f, \tag{11.11}$$

where v is a vector of portfolio weights, R a vector of returns on the risky assets and $\mathbf{1}$ is a vector of ones (so $\mathbf{1}'v$ is the sum of the elements in the v vector). In this case we get

$$v = \Sigma^{-1}(\mu^e/k - \phi S_H), \text{ and} \tag{11.12}$$

$$w = v/(1-\phi), \tag{11.13}$$

where Σ is the covariance matrix of all risky assets (not including the background risk) and S_H is a vector of covariances of the assets with the background risk. In the financial subportfolio, w are the weights on the risky assets and $1 - \mathbf{1}'w$ on the riskfree asset.

Proof. (of (11.12)) The investor solves

$$\max_v v'\mu^e + \phi\mu_H^e + R_f - \frac{k}{2}(v'\Sigma v + \phi^2\sigma_{HH} + 2\phi v'S_H),$$

with first order conditions

$$\mathbf{0} = \mu^e - k(\Sigma v + \phi S_H), \text{ so}$$
$$v = \Sigma^{-1}(\mu^e/k - \phi S_H).$$

■

Remark 11.7 *(Interpreting the hedging terms*) As in the univariate case, the hedging term can be written $-\phi\beta$, where β is from a regression of R^e_H on the vector of investable risky assets (R^e)*

$$R^e_H = \alpha + \beta' R^e + \varepsilon, \text{ with } \beta = \Sigma^{-1} S_H.$$

The portfolio weights in (11.13) will (as long as $\phi S_H \neq 0$) give a return that is off the mean-variance frontier. See Figure 11.2 for an illustration.

Example 11.8 *See Figure 11.2 (left panel) for an illustration of the financial subportfolio w. Notice that the optimal portfolio has lower weights on assets that are positive correlated with the background risk, and vice versa.*

Example 11.9 *(Portfolio choice of a pharmaceutical engineer) 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.*

Remark 11.10 *(Transformed assets*) However, the optimal portfolio w is on the mean-variance frontier of some transformed assets, Z_i. We can rewrite the portfolio return (11.11) as*

$$R_p = \upsilon' Z + (1 - \mathbf{1}'\upsilon) Z_f, \text{ where}$$
$$Z_i = (1 - \phi) R_i + \phi R_H.$$

Notice that all these transformed assets (also Z_f) are risky. The optimal portfolio w from (11.13), that is, set $\upsilon = w$, will be on the mean-variance frontier of Z. See Figure 11.2 (right panel). (The "proof" is that maximizing the objective function (11.3) subject to this new definition of the portfolio return is a traditional mean-variance problem—but in terms of the transformed assets Z.)

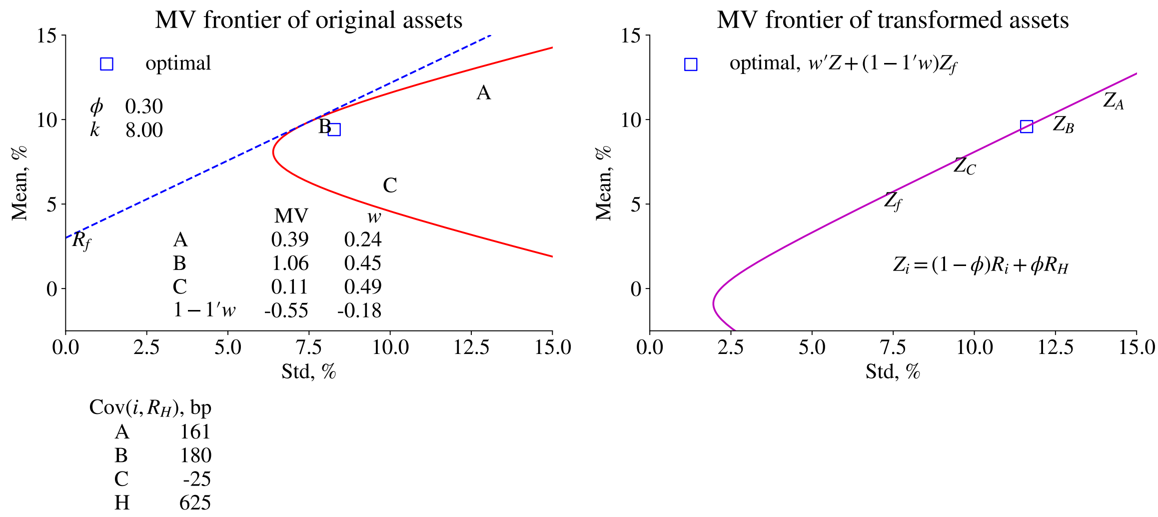


Figure 11.2: Portfolio choice with background risk. The properties of the investable assets (A, B, and C) are shown in Table 11.1.

11.2.3 Asset Pricing Implications of Background Risk I

If the background risk affects portfolio choice for a large fraction of the investors, then it is also likely to influence the (equilibrium) asset pricing. For instance, an asset which provides an effective hedge against background risk will be greatly demanded—and therefore generate low returns.

This leads to an extension of the traditional CAPM expression for expected returns in the form of a *multi-beta model*

$$\mu_i^e = \beta_{im}\mu_m^e + \beta_{iH}\mu_H^e. \tag{11.14}$$

In this expression, β_{im} and β_{iH} are the multiple regression coefficients from the *multi-factor* regression

$$R_i^e = a_i + \beta_{im}R_m^e + \beta_{iH}R_H^e + \varepsilon_i, \tag{11.15}$$

and μ_m^e and μ_H^e are the average excess returns on the two factors. The intercept, a_i, should be zero.

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 (11.14) is that there are two risk factors that influence the required risk premium of asset i: both the market and the background risk.

Example 11.11 *(Multi-factor model) The multiple regression coefficients in (11.15) are*

$$\begin{bmatrix} \beta_{im} \\ \beta_{iH} \end{bmatrix} = \begin{bmatrix} \sigma_{mm} & \sigma_{mH} \\ \sigma_{mH} & \sigma_{HH} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{im} \\ \sigma_{iH} \end{bmatrix}.$$

For instance,

$$\begin{bmatrix} \beta_{im} \\ \beta_{iH} \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 0.8 \\ -0.5 \end{bmatrix} \approx \begin{bmatrix} 1.06 \\ -0.51 \end{bmatrix}.$$

If $\mu_m^e = 0.08$ *and* $\mu_H^e = 0.089$*, the expected excess return of asset* i *should be*

$$\mu_i^e = 1.06 \times 0.08 - 0.51 \times 0.089 = 0.039.$$

Proof. (*of (11.14)) Divide the portfolio weights in (11.12) 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

$$\begin{aligned} \sigma_{pm} &= w_p' \Sigma w_m \\ &= w_p' \Sigma \Sigma^{-1} \left(\mu^e / k - S_H \phi\right) / \left(1 - \phi\right) \\ &= \mu_p^e / \left[k \left(1 - \phi\right)\right] - \sigma_{pH} \phi / \left(1 - \phi\right). \end{aligned}$$

Rewrite as

$$\begin{aligned} \mu_p^e / k &= \left(1 - \phi\right) \sigma_{pm} + \phi \sigma_{pH} \qquad (*) \\ &= \begin{bmatrix} 1 - \phi & \phi \end{bmatrix} \begin{bmatrix} \sigma_{pm} \\ \sigma_{pH} \end{bmatrix} \\ &= \begin{bmatrix} 1 - \phi & \phi \end{bmatrix} \begin{bmatrix} \sigma_{mm} & \sigma_{mH} \\ \sigma_{mH} & \sigma_{HH} \end{bmatrix} \begin{bmatrix} \sigma_{mm} & \sigma_{mH} \\ \sigma_{mH} & \sigma_{HH} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{pm} \\ \sigma_{pH} \end{bmatrix} \\ &= \begin{bmatrix} 1 - \phi & \phi \end{bmatrix} \begin{bmatrix} \sigma_{mm} & \sigma_{mH} \\ \sigma_{mH} & \sigma_{HH} \end{bmatrix} \begin{bmatrix} \beta_{pm} \\ \beta_{pH} \end{bmatrix} \\ &= \begin{bmatrix} \left(1 - \phi\right) \sigma_{mm} + \phi \sigma_{mH} & \left(1 - \phi\right) \sigma_{mH} + \phi \sigma_{HH} \end{bmatrix} \begin{bmatrix} \beta_{pm} \\ \beta_{pH} \end{bmatrix}. \qquad (**) \end{aligned}$$

The third line just multiplies and divides by the covariance matrix. The fourth line follows from the usual definition of regression coefficients, $\beta = \mathrm{Var}(x)^{-1} \mathrm{Cov}(x, y)$.

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

$$\mu^e_m/k = (1-\phi)\,\sigma_{mm} + \phi\sigma_{mH} \text{ and}$$
$$\mu^e_H/k = (1-\phi)\,\sigma_{mH} + \phi\sigma_{HH}.$$

Use these to substitute for the row vector in (**) to get

$$\mu^e_p/k = \begin{bmatrix}\mu^e_m/k & \mu^e_H/k\end{bmatrix}\begin{bmatrix}\beta_{pm}\\ \beta_{pH}\end{bmatrix},$$

which is the same as (11.14). ■

11.2.4 Asset Pricing Implications of Background Risk II: Reinterpretation of CAPM Results

Consider the standard CAPM regression

$$R^e_i = \alpha_i + \beta_i R^e_m + \varepsilon_i, \tag{11.16}$$

where R^e_m is the market excess return. We use time series data to estimate it. It would (in a very large sample) give the traditional beta ($\beta = \sigma_{im}/\sigma_{mm}$), and CAPM suggests that

$$\mu^e_i = \beta_i \mu^e_m, \text{ so } \alpha_i = 0. \tag{11.17}$$

However, this is no longer correct if there is background risk because the estimate of α in (11.16) suffers from an omitted variable bias.

Remark 11.12 *(Omitted variable bias in OLS*) Suppose the correct regression model is $y_t = x_t'\beta + h_t\gamma + u_t$, but we omit the h_t regressor and estimate $y_t = x_t'\beta + \varepsilon_t$ by OLS. It is well known (see, for instance, Verbeek (2004) 5.2), that the OLS estimate converges (as the sample size increases) to $\beta + \theta\gamma$, where θ is from regressing $h_t = x_t'\theta + \eta_t$.*

Remark 11.13 *(Omitted variable bias in CAPM regression*) Assume that the two-factor model (11.15) holds with a zero intercept. Then, applying Remark 11.12 shows that the OLS estimate of α_i from (11.16) is*

$$\hat{\alpha}_i = \hat{\theta}_0 \beta_{iH},$$

where $\hat{\theta}_0$ is the estimate (using time series data) of the intercept in

$$R^e_H = \theta_0 + \theta_1 R^e_m + \eta_t.$$

Together, these two equations suggest that non-zero alphas from CAPM regression may be explained by a combination of (1) *a missing factor ($\beta_{iH} \neq 0$);* (2) *and that factor is not "priced" by the market returns alone ($\theta_0 \neq 0$).*

11.3 Heterogeneous Investors*

This section gives a simple example of a model where the investors have different risk aversions and different beliefs.

Recall the simple MV problem where investor i solves

$$\max_v \mathrm{E}_i\, R_p - \mathrm{Var}_i(R_p) k_i/2, \text{ subject to} \tag{11.18}$$

$$R_p = v R^e_m + R_f. \tag{11.19}$$

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

$$v_i = \frac{1}{k_i} \frac{\mathrm{E}_i\, R^e_m}{\mathrm{Var}_i(R^e_m)}, \tag{11.20}$$

where $\mathrm{E}_i\, R^e_m$ is the investor's expectation of the excess return of the risky asset and $\mathrm{Var}_i(R^e_m)$ the investor's perceived variance.

If all investors have the same initial wealth, then the average (across investors) v_i must be unity—since the riskfree asset is in zero net supply. Suppose there are N investors: the average of (11.20) is

$$1 = \frac{1}{N} \sum\nolimits_{i=1}^{N} \frac{1}{k_i} \frac{\mathrm{E}_i\, R^e_m}{\mathrm{Var}_i(R^e_m)}. \tag{11.21}$$

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, (11.21) can be rearranged as

$$\mathrm{E}\, R^e_m = \tilde{k}\, \mathrm{Var}(R^e_m), \text{ where } \tilde{k} = \frac{1}{\frac{1}{N}\sum_{i=1}^{N} \frac{1}{k_i}}. \tag{11.22}$$

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 11.14 *("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, (11.21) can be rearranged as

$$\mathrm{E}\, R_m^e = \frac{1}{\frac{1}{N}\sum_{i=1}^{N} \frac{1}{k_i \operatorname{Var}_i(R_m^e)}}. \tag{11.23}$$

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, (11.21) can be rearranged as

$$\frac{1}{N}\sum_{i=1}^{N} \mathrm{E}_i\, R_m^e = k \operatorname{Var}(R_m^e). \tag{11.24}$$

Clearly, the average expected excess return is increasing in the risk aversion and variance.

These special cases suggest that, although the general expression (11.21) is complicated, we are unlikely to commit serious errors by sticking to the formulation

$$\mathrm{E}\, R_m^e = k \operatorname{Var}(R_m^e), \tag{11.25}$$

as long as we interpret the components as (close to) averages across investors.

11.4 Joint Portfolio and Savings Choice

11.4.1 Two-Period Problem

The basic *consumption-based* multi-period investment 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

$$\max u(C_1) + \delta \,\mathrm{E}_1\, u(C_2), \text{ subject to} \tag{11.26}$$

$$C_1 + I_1 = W_1 \tag{11.27}$$

$$C_2 + I_2 = \left(1 + R_p\right) I_1 + y_2, \text{ where} \tag{11.28}$$

$$R_p = v_i R_i^e + v_j R_j^e + R_f. \tag{11.29}$$

In equation (11.26) 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 (11.27) 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 (11.28) 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. 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_i and v_j) as well as on the returns on the assets (from holding them from period 1 to period 2).

It is typically difficult to get any closed form solutions for the optimal solution. However, we can gain some insights by studying the first order conditions.

Optimal Investment Level

We first discuss the choice of the *investment level* in period 1, I_1. The composition of the portfolio is discussed later.

It is clear that $I_2 = 0$ since investing in period 2 is just wasting. Use the budget constraints and $I_2 = 0$ to substitute for C_1 and C_2 in (11.26) to get

$$\max u(W_1 - I_1) + \delta \,\mathrm{E}_1\, u[\left(1 + R_p\right) I_1 + y_2]. \tag{11.30}$$

The first order condition for I_1 is that the derivative of (11.30) wrt I_1 is zero

$$-u'(C_1) + \delta \,\mathrm{E}_1 \left[u'(C_2)\left(1 + R_p\right)\right] = 0, \tag{11.31}$$

where $u'(C_t)$ is the marginal utility in period t. (In this expression, the consumption levels are substituted back—in order to facilitate the interpretation.) This expression 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 on the investment.

Example 11.15 *(CRRA and log utility) With a CRRA utility function, $C^{1-\gamma}/(1-\gamma)$, marginal utility is $C^{-\gamma}$. With a logarithmic utility function, marginal utility is $1/C$.*

We can also rewrite (11.31) as

$$\mathrm{E}_1\left[\frac{\delta u'(C_2)}{u'(C_1)}\left(1+R_p\right)\right]=1. \tag{11.32}$$

For instance, with logarithmic utility (11.32) is

$$\mathrm{E}_1\left[\delta\frac{C_1}{C_2}\left(1+R_p\right)\right]=1 \text{ (with log utility)}. \tag{11.33}$$

This expression suggests that when $1+R_p$ tends to be high, then C_1/C_2 will tend to be low—or else the expected product will not be equal to one. The economic mechanism is that when R_p is (expected to be) high, then it is worthwhile to invest (save rather than consume). This is a key issue in macroeconomics, but not the focus in portfolio choice models.

Example 11.16 *(Planned consumption profile when the portfolio is riskfree) As a special case, suppose the investor holds only riskfree assets. The portfolio return is then R_f, which is non-random. The log utility case (11.33) can then be written*

$$\mathrm{E}_1\frac{C_1}{C_2}=\frac{1}{\delta}\frac{1}{1+R_f}.$$

With $\delta=1/1.01$, the planned consumption profile at different interest rates would be

R_f	$\mathrm{E}_1 C_1/C_2$
0.01	$1.01\frac{1}{1.01}=1$
0.05	$1.01\frac{1}{1.05}=0.96$

Optimal Portfolio Choice

We now consider the *portfolio choice*, that is, the portfolio weights v_i and v_j. We use the definition of the portfolio return (11.29) in the objective function (11.30) to get

$$\max u\left(W_1-I_1\right)+\delta\,\mathrm{E}_1\,u\left[\left(1+v_iR_i^e+v_jR_j^e+R_f\right)I_1+y_2\right]. \tag{11.34}$$

The first order conditions for v_i and v_j are

$$\mathrm{E}_1\, u'(C_2) R_i^e = 0 \text{ and} \tag{11.35}$$

$$\mathrm{E}_1\, u'(C_2) R_j^e = 0. \tag{11.36}$$

This says that both excess returns should be "orthogonal" to marginal utility.

To solve for the decision variables (I_1, v_i, v_j) we should use the budget restrictions to substitute for C_1 and C_2 in the first order conditions (11.32), (11.35) and (11.36)—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.

Remark 11.17 *(Rewriting* $\mathrm{E}\, xy = 0$*) Recall that, by definition,* $\mathrm{Cov}(x, y) = \mathrm{E}\, xy - \mathrm{E}\, x \times \mathrm{E}\, y$*, or* $\mathrm{E}\, xy = \mathrm{Cov}(x, y) + \mathrm{E}\, x \times \mathrm{E}\, y$*.* $\mathrm{E}\, xy = 0$ *therefore implies that* $\mathrm{E}\, y = \mathrm{Cov}(-x, y)/\,\mathrm{E}\, x$*.*

The first order conditions still contain some useful information. Use Remark 11.17 to write (11.35) as

$$\mathrm{E}\, R_i^e = \frac{\mathrm{Cov}[-u'(C_2), R_i^e]}{\mathrm{E}\, u'(C_2)}, \tag{11.37}$$

and similarly for (11.36). This expression says that (in equilibrium) asset i 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 (pro-cyclical assets) 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. In short, *procyclical assets are risky*—and must therefore give high expected returns.

Remark 11.18 *(*Linearizing* $u'(C)$*) A first-order Taylor approximation of marginal utility around* $\bar{C}$ *is* $u'(C) \approx u'(\bar{C}) + u''(\bar{C})(C - \bar{C})$*. Using a Taylor approximation of* $u'(C_2)$*, the covariance in (11.37) can be written* $\mathrm{Cov}[-u'(C_2), R_i^e] \approx A\, \mathrm{Cov}(C_2, R_i^e)$*,where* A *is a positive constant (equal to* $-u''(\bar{C})$*, which is positive since the utility function is concave).*

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

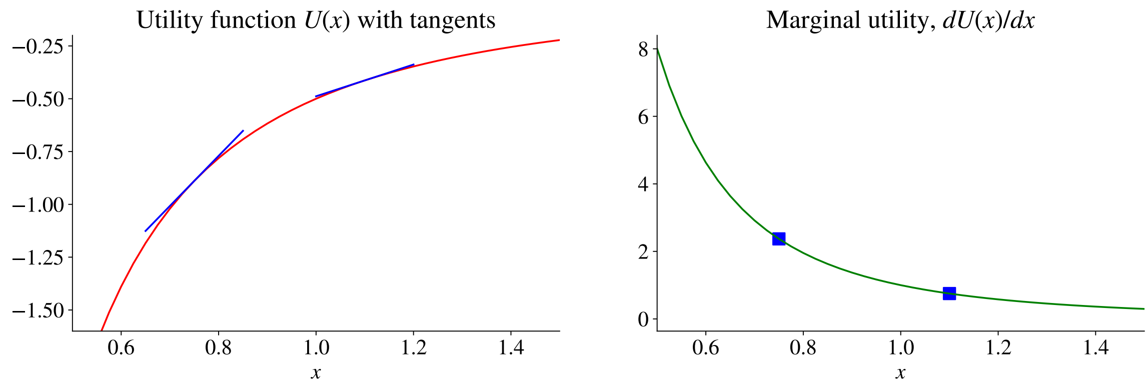


Figure 11.3: Utility function

function is

$$u(C_1) + \delta \,\mathrm{E}_1\, u(C_2) + \delta^2 \,\mathrm{E}_1\, u(C_3) + \ldots \delta^{T-1} \,\mathrm{E}_1\, u(C_T). \tag{11.38}$$

11.4.2 From a Consumption-Based Model to CAPM

Suppose marginal utility is an affine function of the market excess return

$$u'(C_2) = a - bR^e_m, \text{ with } b > 0. \tag{11.39}$$

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 (11.37) as

$$\mathrm{E}\, R^e_i = b\frac{\mathrm{Cov}(R^e_m, R^e_i)}{\mathrm{E}(a - bR^e_m)}. \tag{11.40}$$

We can, of course, apply this expression to the market excess return (instead of asset 1) to get

$$\mathrm{E}\, R^e_m = b\frac{\mathrm{Var}(R^e_m)}{\mathrm{E}(a - bR^e_m)}. \tag{11.41}$$

Combine (11.41) in (11.40) to to get

$$\mathrm{E}\, R^e_i = \frac{\mathrm{Cov}(R^e_m, R^e_i)}{\mathrm{Var}(R^e_m)}\,\mathrm{E}\, R^e_m, \tag{11.42}$$

which is the beta representation of CAPM. This means that CAPM is consistent with (some) multi-period utility based portfolio choice models.

11.4.3 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 (y) and interest rates (r)

$$-u'(C_2) = ay + br. \tag{11.43}$$

Such a formulation makes a lot of sense in most macro models—at least as an approximation. It is then possible to write (11.37) as

$$\mathrm{E}\, R_i^e = \frac{a\,\mathrm{Cov}(y, R_i^e) + b\,\mathrm{Cov}(r, R_i^e)}{-\mathrm{E}(ay + br)}. \tag{11.44}$$

This, in turn, is easily put in the form of (11.2), where the risk premium on asset 1 depends on the betas against GDP and the interest rate. (See the proof of (11.14) for an idea of how to construct this beta representation.)

11.5 Testing Multi-Factors Models

Provided all factors are excess returns, we can test a multi-factor model by testing whether $\alpha = 0$ in the regression

$$R_{it}^e = \alpha + b_{io} R_{ot}^e + b_{ip} R_{pt}^e + ... + \varepsilon_{it}. \tag{11.45}$$

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

$$\frac{\hat{\alpha}_i}{\mathrm{Std}(\hat{\alpha}_i)} \xrightarrow{d} N(0, 1) \text{ under } H_0 : \alpha_i = 0. \tag{11.46}$$

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). 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 11.19 *(Fama-French factors) Fama and French (1993) use three factors: the market excess 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 a high ratio of book value to market value minus the return on a portfolio with a low ratio (HML). All three* *are excess returns (although only the first is in excess of a riskfree return), since they are long-short portfolios. He and Ng (1994) try to relate these factors to macroeconomic series.*

Remark 11.20 *(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, $R_i^e - R_j^e$). 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 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 11.4 shows some results for the Fama-French model on US industry portfolios and Figures 11.5–11.6 on the 25 Fama-French portfolios.

11.6 Appendix: Extra Material

11.6.1 The Arbitrage Pricing Model*

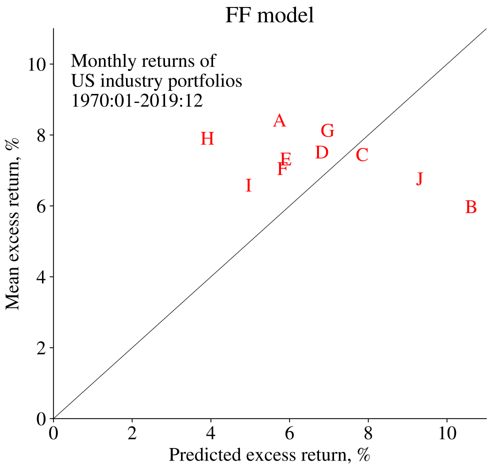
Reference: Ross (1976)

The *first assumption* of the Arbitrage Pricing Theory (APT) is that the return of asset i can be described as

$$R_{it} = a_i + \beta_i f_t + \varepsilon_{i,t}, \text{ where} \quad (11.47)$$
$$\mathrm{E}\,\varepsilon_{it} = 0, \mathrm{Cov}(\varepsilon_{it}, f_t) = \mathrm{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0.$$

In this particular formulation there is only one factor, f_t, but the APT allows for more factors. Notice that (11.47) assumes that any correlation of two assets (i and j) is due to movements in f_t—the residuals are assumed to be uncorrelated.

The *second assumption* of APT is that 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



	α	t-stat	σ
A (NoDur)	2.62	2.16	8.43
B (Durbl)	-4.69	-2.73	11.94
C (Manuf)	-0.45	-0.53	5.94
D (Enrgy)	0.64	0.31	14.48
E (HiTec)	1.36	0.97	9.75
F (Telcm)	1.18	0.79	10.42
G (Shops)	1.11	0.85	9.04
H (Hlth)	3.94	2.65	10.33
I (Utils)	1.57	1.01	10.86
J (Other)	-2.60	-3.26	5.55

Fama-French model
Factors: US market, SMB (size), and HML (book-to-market)
Predicted excess return: $\beta_m R_m^e + \beta_{SMB} R_{SMB} + \beta_{HML} R_{HML}$

α and σ (std of residual) are in annualized %

p-val for testing if all $\alpha_i = 0$: 0.0

Figure 11.4: Fama-French regressions on US industry indices

portfolio (buy some, sell some) that has a zero exposure 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 $\Sigma_{i=1}^N v_i = 0$ (zero cost), $\Sigma_{i=1}^N v_i \beta_i = 0$ (zero exposure to the factor) and $\Sigma_{i=1}^N v_i^2 \operatorname{Var}(\varepsilon_{i,t}) \approx 0$ (well diversified). The requirement that the portfolio is non-trivial means that at least some $v_j \neq 0$. This riskfree portfolio has a zero cost, so it must have a return of zero (and thus also an expected return of zero) or otherwise there are arbitrage opportunities.

Together, these assumptions imply that for (every asset) we have

$$\mathrm{E}\, R_{it} = R_f + \beta_i \lambda, \tag{11.48}$$

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 11.21 *(APT with three assets) Suppose there are three well-diversified portfo-*

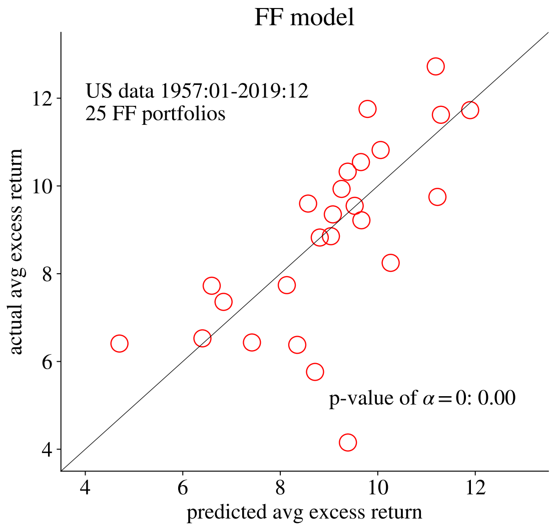


Figure 11.5: FF, FF portfolios

lios (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 11.22 *(APT with two assets) Example 11.21 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 $v_1 \neq 0$).*

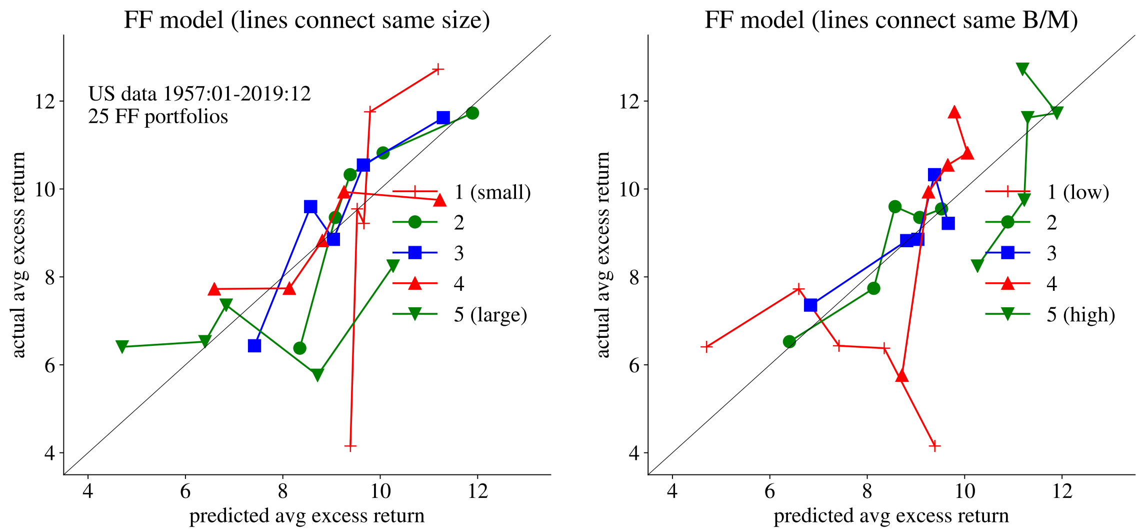


Figure 11.6: FF, FF portfolios

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.

11.6.2 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, $\text{Cov}(R_m, R_z) = 0$. (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

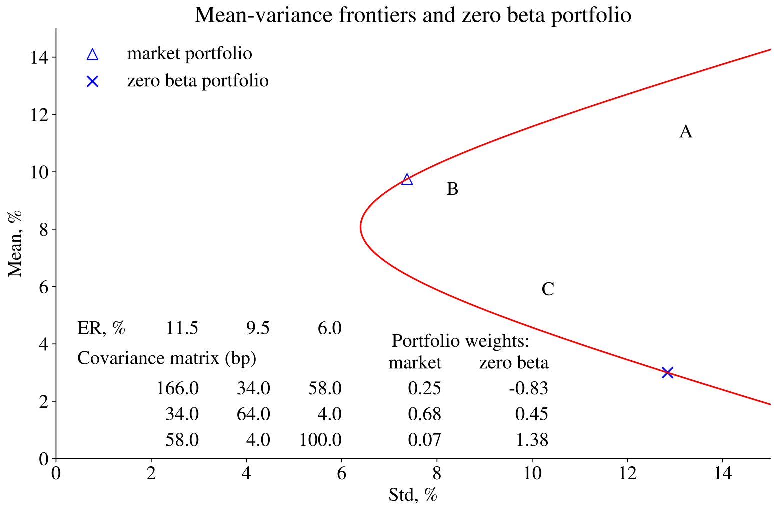


Figure 11.7: Zero-beta model

sign compared to the market portfolio (contributing to a negative covariance). Together, this gives a zero covariance.

See Figure 11.7 for an illustration.

The main result is then the "zero-beta" CAPM

$$\mathrm{E}(R_i - R_z) = \beta_i \,\mathrm{E}(R_m - R_z). \tag{11.49}$$

Proof. (*of (11.49)) An investor (with initial wealth equal to unity) chooses the portfolio weights (v_i) to maximize

$$\begin{aligned}\mathrm{E}\, U(R_p) &= \mathrm{E}\, R_p - \frac{k}{2}\mathrm{Var}(R_p), \text{ where}\\ R_p &= v_1 R_1 + v_2 R_2 \text{ and } v_1 + v_2 = 1,\end{aligned}$$

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

$$\sigma_{1m} = \mathrm{Cov}(R_1, \underbrace{v_1 R_1 + v_2 R_2}_{R_m}) = v_1\sigma_{11} + v_2\sigma_{12},$$

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

$$\begin{aligned} v_1\sigma_{1m} + v_2\sigma_{2m} &= v_1\,\mathrm{Cov}(R_1, v_1R_1 + v_2R_2) + v_2\,\mathrm{Cov}(R_2, v_1R_1 + v_2R_2) \\ &= \mathrm{Cov}(v_1R_1 + v_2R_2, v_1R_1 + v_2R_2) \\ &= \mathrm{Var}(R_m). \end{aligned}$$

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 (11.49). ■

Chapter 12

Investment for the Long Run

Reference: Campbell and Viceira (2002)

12.1 Time Diversification

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 12.1.

12.1.1 Long-Run Return as a Sum of Short-Run Returns

This section shows how a long-run return can be expressed in terms of many short-run returns—and how common approximations work.

The gross return on a q-period investment made in period 0 can be written

$$1 + Z_q = (1 + R_1)(1 + R_2)...(1 + R_q), \tag{12.1}$$

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

$$z_q = r_1 + r_2 + \ldots + r_q, \tag{12.2}$$

where $z_q = \ln(1+Z_q)$ and $r_t = \ln(1+R_t)$. Notice that if R is small, then $\ln(1+R) \approx R$. We use r_t^e to denote the excess long return, $r_t^e = r_t - r_f$, where $r_f = \ln(1 + R_f)$, and similarly for z_q^e.

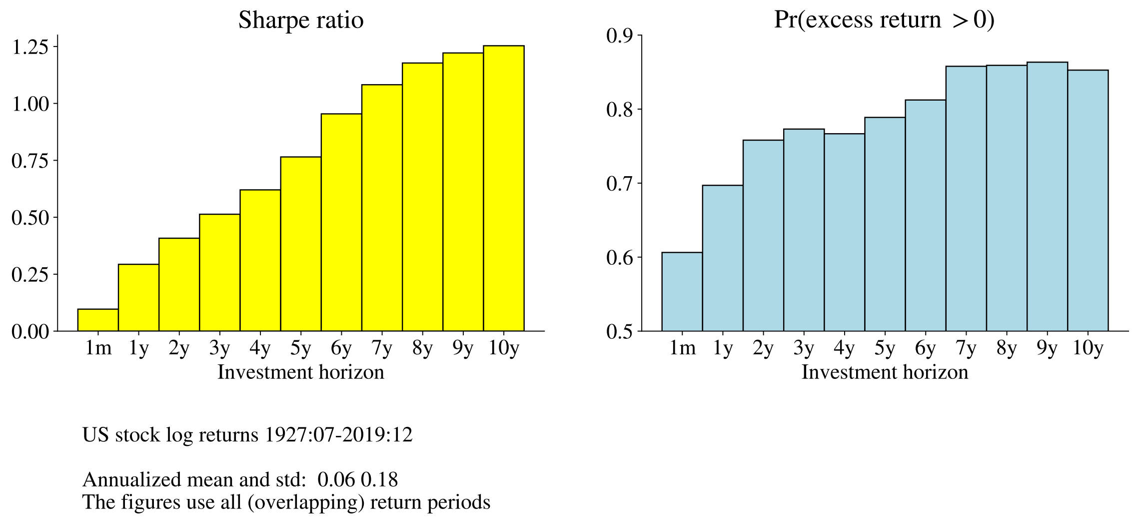


Figure 12.1: Empirical evidence on SR and probability of excess return > 0

It is sometimes convenient to approximate the q-period return Z_q (not the log) as

$$
\begin{aligned}
Z_q &= (R_1 + 1)(R_2 + 1)\dots(R_q + 1) - 1 \\
&\approx R_1 + R_2 + \dots + R_q. \qquad (12.3)
\end{aligned}
$$

This approximation often works well, unless there are very many periods and/or there are some extreme one-period returns.

Example 12.1 *(The quality of the approximation of the q-period return 2) Consider the following table of net returns*

	Portfolio A	*Portfolio B*
Year 1	5%	20%
Year 2	−5%	−35%
Year 3	5%	25%
Total return over 3 years	4.7%	−2.5%
Average net return	1.67%	3.33%
Average log return	1.54%	−0.84%

Example 12.2 *(The quality of the approximation of the q-period return 2) If $R_1 = 0.9$ and $R_2 = -0.9$ (indeed very extreme returns), then the two-period net return is*

$$Z_2 = (1 + 0.9)(1 - 0.9) - 1 = -0.81$$

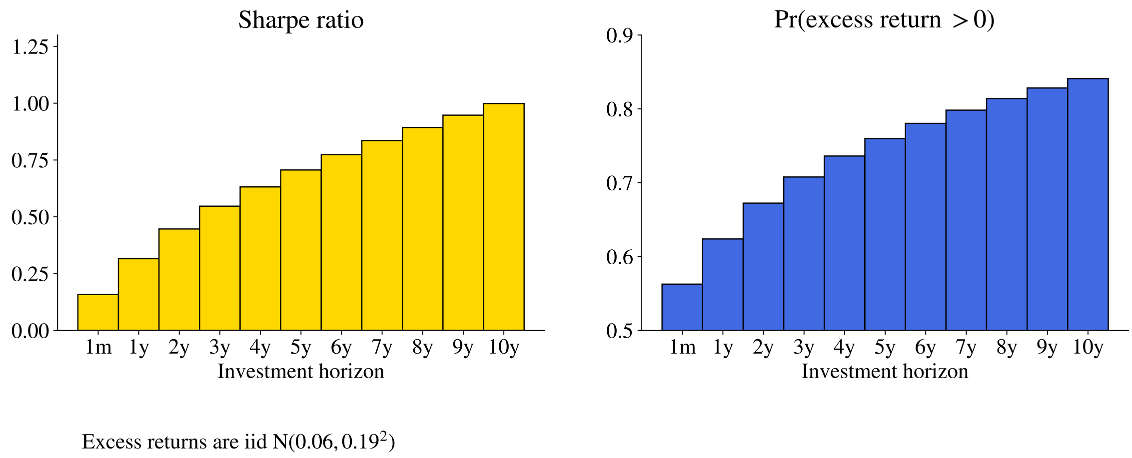


Figure 12.2: SR and probability of excess return > 0, iid returns

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.

Remark 12.3 *(Geometric and arithmetic average returns*) The average log return,* $\Sigma_{t=1}^{q} r_t/q$*, is closely related to the geometric mean return. To see that, notice that a geometric mean return* $\tilde{r}$ *satisfies* $(1 + \tilde{r})^q = 1 + Z_q$*. Take logs and divide by q to get* $\Sigma_{t=1}^{q} r_t/q$*.*

12.1.2 Increasing Sharpe Ratios

This section demonstrates that, with iid (unpredictable) returns, the expected return and variance of a portfolio both grow linearly with the investment horizon, so Sharpe ratios (expected excess return divided by the standard deviation) increase with the square root of horizon.

Let z_q be the log return on a q-period investment. *If log returns are iid*, the Sharpe ratio of z_q is

$$SR(z_q) = \sqrt{q}SR(r), \tag{12.4}$$

where $SR(r)$ is the Sharpe ratio of the *one-period* log return. (Time subscripts are suppressed to keep the notation simple.) $SR(z_q)$ is clearly increasing with the horizon, q, provided $SR(r) > 0$.

Proof. (of (12.4)) The q-period log return is as in (12.2). If one-period excess log returns are iid with mean μ^e and variance σ^2, then the mean and variance of the q-period excess log returns are

$$\begin{aligned} \mathrm{E}\, z_q^e &= q\mu^e, \\ \mathrm{Var}(z_q) &= q\sigma^2. \end{aligned}$$

Equation (12.3) shows that we get approximately the same result for Z_q. ■

12.1.3 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 log returns are normally distributed (not necessarily iid). Then, we have

$$\Pr(z_q^e > 0) = \Phi[SR(z_q)], \tag{12.5}$$

where z_q^e is the excess log return on a q-period investment and $\Phi()$ is the cumulative distribution function of a standard normal variable, $N(0, 1)$. See Figure 12.2 for an illustration.

Proof. (of (12.5)) By standard manipulations we have

$$\begin{aligned} \Pr(z_q^e \leq 0) &= \Pr\left(\frac{z_q^e - \mathrm{E}\, z_q^e}{\mathrm{Std}(z_q^e)} \leq \frac{-\mathrm{E}\, z_q^e}{\mathrm{Std}(z_q^e)}\right) \\ &= \Phi\left(\frac{-\mathrm{E}\, z_q^e}{\mathrm{Std}(z_q^e)}\right). \end{aligned}$$

Clearly, $\Pr(z_q^e > 0) = 1 - \Pr(z_q^e \leq 0)$. Use the fact that $\Phi(x) + \Phi(-x) = 1$ (since the standard normal distribution is symmetric around zero) to get (12.5). ■

12.1.4 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 necessarily mean that the risky asset is considered to be safer for a long-run investor. The reason is, of course, that we

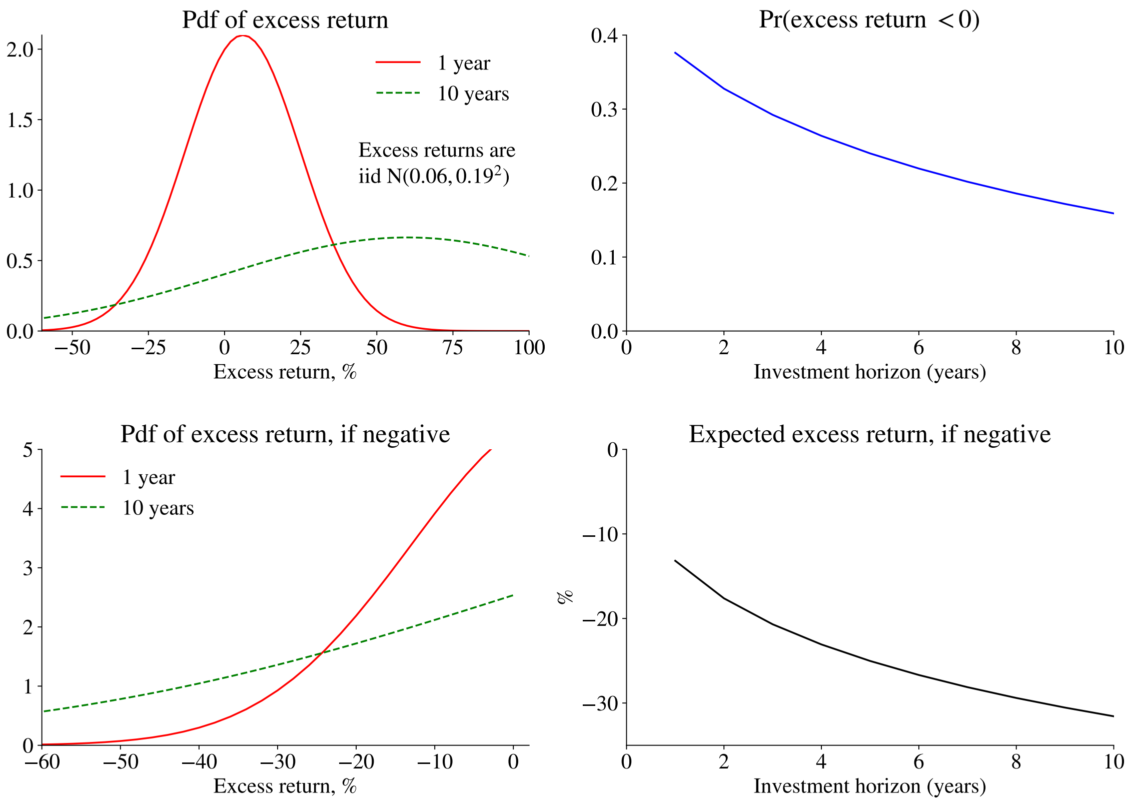


Figure 12.3: Time diversification, normally distributed returns

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 12.3 for an illustration.

Remark 12.4 *(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* $\mathrm{E}(x|x \leq 0) = \mu - \sigma\phi(-\mu/\sigma)/\Phi(-\mu/\sigma)$.

To say more about how the investment horizon affects the portfolio weights, we need to be more precise about the preferences, that is, how risks and opportunities are compared. As a benchmark, consider a mean-variance investor who will choose a portfolio for q periods. With one risky asset (the tangency portfolio, perhaps) and a riskfree asset, the q-period return in $vZ_q + (1-v)qR_f = vZ_q^e + qR_f$. In this case, we chose to work with net returns (rather than log returns), since they are linear in portfolio weights. (To obtain similar analytical results for log returns, we could instead make a Taylor approximation of the log portfolio return to make it (approximately) linear in the portfolio weights. See Campbell and Viceira (2002).)

The optimization problem for a q-period investment (chosen in period 0) is then

$$\max_v \ v \, \mathrm{E}_0 \, Z_q^e + qR_f - \frac{k}{2} v^2 \, \mathrm{Var}_0(Z_q^e) \tag{12.6}$$

where R_f is the per-period riskfree rate. We assume that the investor has to choose a portfolio for the entire investment period (no rebalancing). The expectation and covariance carry time subscripts to indicate that they are conditional moments (that is, based on the information when the portfolio is chosen).

Remark 12.5 ***Example 12.6*** *(Conditional moments*) Suppose* $x_{t+1} = \rho x_t + \varepsilon_{t+1}$ *and* $y_{t+1} = \pi y_t + v_{t+1}$*, where* ε_{t+1} *and* v_{t+1} *are unpredictable and have constant variances (and covariance). Then,* $\mathrm{E}_t \, x_{t+1} = \rho x_t$*,* $\mathrm{Var}_t(x_{t+1}) = \mathrm{Var}(\varepsilon_{t+1})$ *and* $\mathrm{Cov}_t(x_{t+1}, y_{t+1}) = \mathrm{Cov}(\varepsilon_{t+1}, v_{t+1})$*. If* $p = 0$ *and* $\pi = 0$*, then these conditional moments are the same as the unconditional moments.*

The solution is that the portfolio weight on the risky asset is

$$v = \frac{1}{k} \frac{\mathrm{E}_0 \, Z_q^e}{\mathrm{Var}_0(Z_q^e)}, \tag{12.7}$$

with $1 - v$ in the riskfree asset.

iid Returns

With iid returns, both the mean and the variance scale linearly with the investment horizon and the conditional moments are the same as the unconditional ones (so $\mathrm{E}_0 \, Z_q^e = q \, \mathrm{E} \, R_1^e$ and $\mathrm{Var}_0(Z_q^e) = q \, \mathrm{Var}(R_1^e)$). Using this in (12.7) gives

$$v = \frac{1}{k} \frac{\mathrm{E} \, R_1^e}{\mathrm{Var}(R_1^e)}, \text{ if iid returns.} \tag{12.8}$$

With MV preferences and *iid returns, the investment horizon does not matter for the portfolio choice*. Notice the difference to the Sharpe ratio and the probability of beating the riskfree asset which increase with the horizon. In short, MV analysis suggests that comparing Sharpe ratios across investment horizons is not very informative.

Autocorrelated Returns

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 weight for a one-period investor on the risky asset is then

$$v = \frac{1}{k}\frac{\mathrm{E}_0\, R_1^e}{\mathrm{Var}_0(R_1^e)}. \tag{12.9}$$

For a two-period investor the portfolio weight is instead

$$v = \frac{1}{k}\frac{\mathrm{E}_0(R_1^e + R_2^e)}{\mathrm{Var}_0(R_1^e) + \mathrm{Var}_0(R_2^e) + 2\,\mathrm{Cov}_0(R_1^e, R_2^e)}. \tag{12.10}$$

Notice 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* since extreme movements will be partially "averaged out" over time. Clearly, momentum would give the opposite effect.

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.

Proof. (of (12.10)) The first order condition of (12.6) is

$$\begin{aligned} 0 &= \mathrm{E}\, Z_q^e - kv\,\mathrm{Var}(Z_q^e) \text{ or} \\ v &= \frac{1}{k}\frac{\mathrm{E}\, Z_q^e}{\mathrm{Var}(Z_q^e)}. \end{aligned}$$

For the two-period horizon, we use the approximation that $Z_2^e = R_1^e + R_2^e$, so its expected value is $\mathrm{E}_0(R_1^e + R_2^e)$ and its variance$\mathrm{Var}_0(R_1^e + R_2^e) = \mathrm{Var}_0(R_1^e) + \mathrm{Var}_0(R_2^e) + 2\,\mathrm{Cov}_0(R_1^e, R_2^e)$. ■

Example 12.7 *(AR(1) process for returns*) Suppose the excess returns follow an AR(1) process*

$$R_{t+1}^e = \mu(1-\rho) + \rho R_t^e + \varepsilon_{t+1} \text{ with } \sigma^2 = \mathrm{Var}(\varepsilon_{t+1}).$$

We can therefore write R_{t+2}^e *as*

$$\begin{aligned} R_{t+2}^e &= \mu(1-\rho) + \rho R_{t+1}^e + \varepsilon_{t+2} \\ &= \mu(1-\rho^2) + \rho^2 R_t^e + \rho\varepsilon_{t+1} + \varepsilon_{t+2}. \end{aligned}$$

The conditional moments are then easily seen to be

$$
\begin{aligned}
\mathrm{E}_0 R_1^e &= \mu(1-\rho) + \rho R_0^e, \\
\mathrm{E}_0 R_2^e &= \mu(1-\rho^2) + \rho^2 R_0^e, \\
\mathrm{Var}_0(R_1^e) &= \sigma^2 \\
\mathrm{Var}_0(R_2^e) &= (1+\rho^2)\sigma^2 \\
\mathrm{Cov}_0(R_1^e, R_2^e) &= \rho\sigma^2.
\end{aligned}
$$

If the initial return is at the mean, $R_0^e = \mu$, then the forecasted return is μ for all future periods, which gives the portfolio weights

$$
\begin{aligned}
v &= \frac{1}{k}\frac{\mu}{\sigma^2} \text{ (for one period)} \\
v &= \frac{1}{k}\frac{\mu}{\sigma^2}\frac{2}{(2+\rho^2+2\rho)} \text{ (for two periods).}
\end{aligned}
$$

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 ratio of the two portfolio weights is around $(1.1, 1, 0.9)$.

12.2 Long-Run Portfolio Choice with a Logarithmic Utility Function (with Rebalancing)

We now (realistically) allow investors to *rebalance* their portfolios in every period. Consider a logarithmic utility function. The objective in period 0 is then

$$\max \mathrm{E}_0 \ln W_q = \max(\ln W_0 + \mathrm{E}_0 r_1 + \mathrm{E}_0 r_2 + \ldots + \mathrm{E}_0 r_q), \tag{12.11}$$

where r_t is the log return, $r_t = \ln(1 + R_t)$ where R_t is a net return. Recall that $\ln W_q = \ln W_0 + r_1 + r_2 + \ldots + r_q$.

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 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 (12.15), but with a logarithmic utility function.

Finding the portfolio weights that maximize $\mathrm{E}_0 r_1$ is not trivial, because of the non-

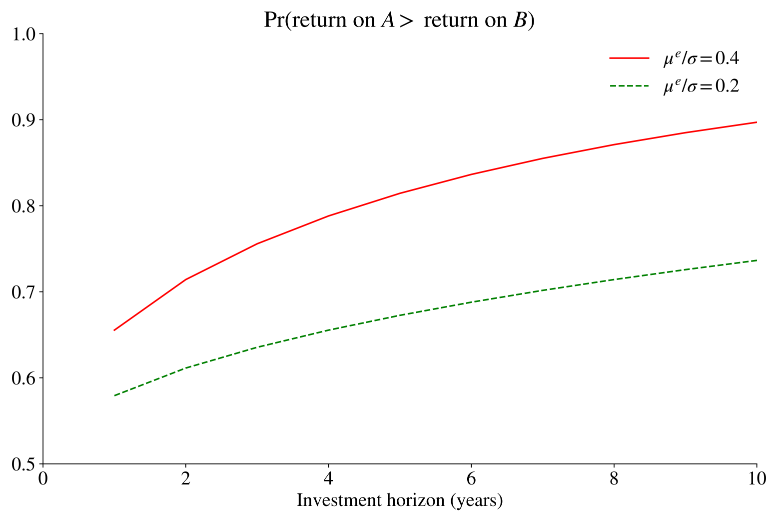


Figure 12.4: The probability of outperforming another portfolio

linearity. The problem is to solve

$$\max \mathrm{E} \ln(w' R^e + R_f), \text{ with foc} \tag{12.12}$$
$$\mathrm{E} \frac{R_i^e}{w' R^e + R_f} = 0 \text{ for } i = 1...n,$$

which involves first calculating the expectations and then solving n non-linear equations for the unknown w vector. Irrespective of these practical issues, the conclusions about (12.11) are right. In particular, all investors with logarithmic utility functions choose the sample portfolios—irrespective of their investment horizons.

Remark 12.8 *(Growth optimal portfolio*) The portfolio that maximizes* $\mathrm{E}_0 \, r_1$ *is often called the "growth optimal" portfolio, since it aims to grow the portfolio value as much as possible. To illustrate that it works, consider the case of iid returns. In this case, let* z_q^e *in (12.5) represent the log return of a portfolio A minus the log return of portfolio B over q periods. The equation then shows that the probability of A outperforming B equals* $\Phi[SR(z_q)]$ *where* $SR(z_q)$ *should be interpreted as* $q(\mathrm{E}\, r_q^A - \mathrm{E}\, r_q^B)/[\sqrt{q}\, \mathrm{Std}(r_q^A - r_q^B)]$. *Clearly, if A has the highest* $\mathrm{E}\, r$ *of all portfolios, then the probability that it outperforms any other portfolio is increasing with time. See Figure 12.4 for an illustration.*

Remark 12.9 *(Maximizing the Geometric Mean Return*) The growth-optimal portfolio is often said to maximize the geometric mean return. To see this, notice from Example*

12.3 that the geometric mean return is an increasing function of the average log return. Maximizing one of them means maximizing the other.

12.3 More General Utility Functions and Rebalancing

We will now take a look at *more general optimization problems*. Assume that the objective is to maximize

$$\mathrm{E}_0\, u(W_q), \tag{12.13}$$

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 12.1.2 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.

12.3.1 CRRA Utility Function

Suppose the utility function has constant relative risk aversion, so the objective in period 0 is

$$\max \mathrm{E}_0\, W_q^{1-\gamma}/(1-\gamma). \tag{12.14}$$

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.

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

$$\max C_0^{1-\gamma}/(1-\gamma) + \delta \operatorname{E}_0 C_1^{1-\gamma}/(1-\gamma) + \ldots + \delta^q \operatorname{E}_0 C_q^{1-\gamma}/(1-\gamma). \qquad (12.15)$$

The basic mechanism is that the optimal consumption/wealth ratio turns out to be constant.

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 only counter example to this is when $\gamma = 1$, that is, with logarithmic utility. It is a very special case.

Chapter 13

Efficient Markets

Reference (medium): Elton, Gruber, Brown, and Goetzmann (2014) 17 (efficient markets) and 19 (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.

13.1 The Efficient Market Hypothesis

The efficient market hypothesis (EMH) says that it is very *hard to predict future asset returns*. If this is true (evidence is discussed later), then active management (security analysis, market timing) is useless and costly (management fees, trading costs). Instead, it makes more sense to apply a passive approach that meets individual requirements (diversification, hedging background risk, appropriate risk level, etc). The practical implications are thus very significant.

13.1.1 Defining Expected Returns

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 net return (R_{t+1}, like 0.05) of holding an asset with dividends (per current share), D_{t+1}, between t and $t+1$ is then defined as

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} - 1. \tag{13.1}$$

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 13.1 *(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 (13.1) based on the information set in t

$$\mathrm{E}_t\, R_{t+1} = \frac{\mathrm{E}_t\, P_{t+1} + \mathrm{E}_t\, D_{t+1}}{P_t} - 1. \tag{13.2}$$

This formulation is only a definition, but it will help us organize the discussion of how asset prices are determined.

13.1.2 Different Versions of the Efficient Market Hypothesis

The efficient market hypothesis (EMH) 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.

The basic idea of the EMH is illustrated in Figure 13.1, which shows the asset price path around an event. If this is the typical (average) path, then both the overreaction and underreaction paths imply that returns are forecastable (after the event). For instance, after the event the overreaction path shows mean reversion of the price (so it would be profitable to short sell this asset just after the event), while the underreaction path shows the opposite. The EMH suggests that such (easy) opportunities do not exist (at least not systematically), so the (average) price path must be the solid line.

However, a precise formulation of the EMH needs to specify two things. First, what type of information is used in making those forecasts? Is it price and trading volume data (the weak form of the EMH), all public information (the semi-strong form), or perhaps all public and private information (the strong form)? Most modern analysis is focused on the weak or semi-strong forms (as private information is likely to have predictive power). Second, what is supposed to be unforecastable? Is it price changes, returns, or excess returns? This is discussion in some detail below.

If price changes are unforecastable, then $\mathrm{E}_t\, P_{t+1} - P_t$ equals a constant. Typically,

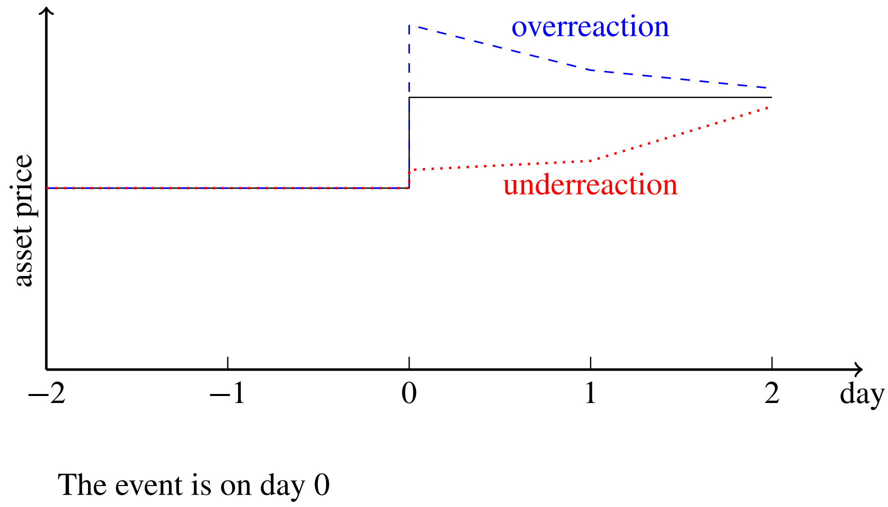


Figure 13.1: Asset price path around an event

this constant is taken to be zero. Use $\mathrm{E}_t\, P_{t+1} = P_t$ in (13.2) to get

$$\mathrm{E}_t\, R_{t+1} = \frac{\mathrm{E}_t\, D_{t+1}}{P_t}. \tag{13.3}$$

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. 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 (13.2) gives $\mathrm{E}_t\, R_{t+1} = 0$, which seems implausible for long investment horizons—although it is probably a reasonable approximation for short horizons (a week or less).

If returns are unforecastable, then $\mathrm{E}_t\, R_{t+1} = R$ (a constant). The main problem with this formulation 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. This is a reasonable null hypothesis, which will be used in these notes.

Rejection of the EMH can have different sources: changes in risk or in risk aversion (both "rational" reasons) or in inefficiencies. It is typically very hard to disentangle these possible sources.

13.2 Autocorrelations and Autoregressions

Autocorrelations and autoregressions are tools for studying whether past and current returns can predict future returns (typically of the same asset).

13.2.1 Autocorrelation Coefficients

The autocovariances of the R_t process can be estimated as

$$\hat{\gamma}_s = \frac{1}{T} \sum_{t=1+s}^{T} \left(R_t - \bar{R}\right)\left(R_{t-s} - \bar{R}\right), \text{ with} \tag{13.4}$$

$$\bar{R} = \frac{1}{T} \sum_{t=1}^{T} R_t. \tag{13.5}$$

(We typically divide by T in (13.4) even if we have only $T-s$ full observations to estimate γ_s from.) Autocorrelations are then estimated as

$$\hat{\rho}_s = \hat{\gamma}_s / \hat{\gamma}_0. \tag{13.6}$$

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). When the true autocorrelations are all zero (not ρ_0, of course), then for any lag s different from zero

$$\sqrt{T}\hat{\rho}_s \to^d N(0,1), \tag{13.7}$$

so $\sqrt{T}\hat{\rho}_s$ can be used as a t-stat. See Figures 13.2–13.3.

Example 13.2 *(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 -1.64 and another 5% above 1.64, we can reject the null hypothesis at the 10% level if* $\sqrt{T}|\hat{\rho}_1| > 1.64$.

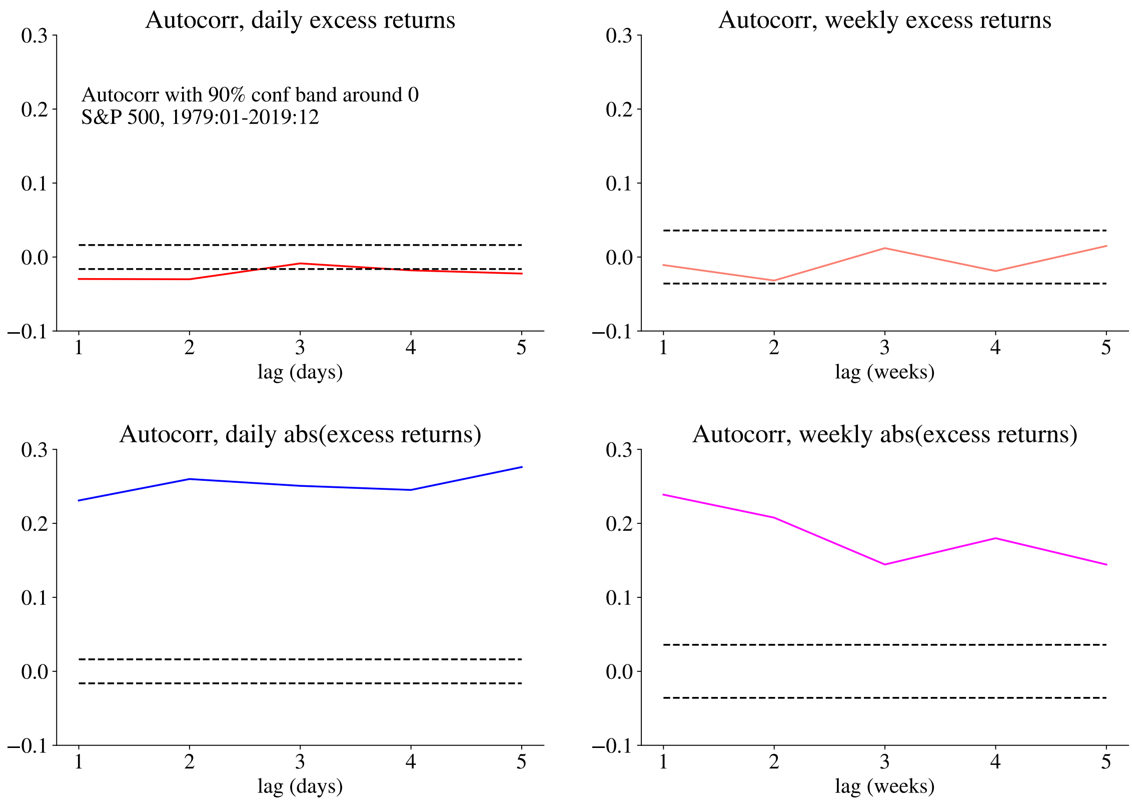


Figure 13.2: Predictability of US stock returns

13.2.2 Autoregressions

An alternative way of testing autocorrelations is to estimate an AR model

$$R_t = c + a_1 R_{t-1} + a_2 R_{t-2} + ... + a_p R_{t-p} + \varepsilon_t, \tag{13.8}$$

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 is easy to implement (and the difference is not large). See Table 13.1 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

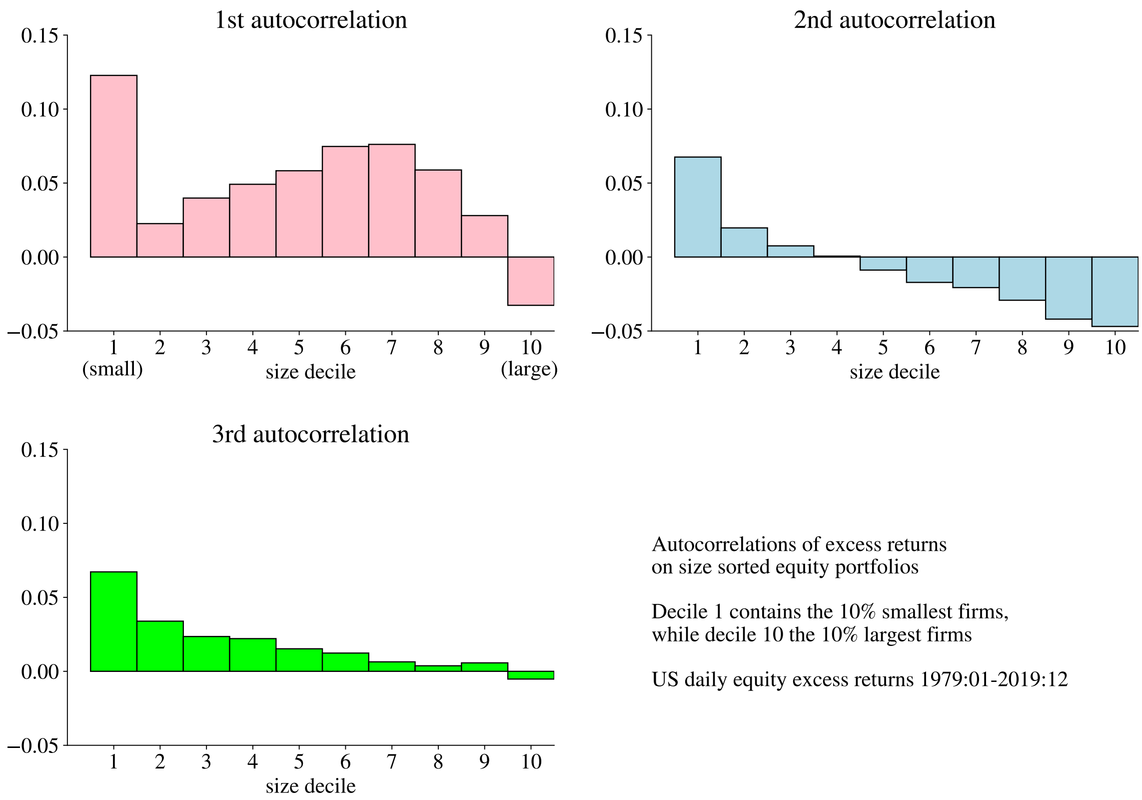


Figure 13.3: Predictability of US stock returns, size deciles

different depending on the sign of last period's return

$$R_t = \alpha + \beta Q_{t-1} R_{t-1} + \gamma(1 - Q_{t-1}) R_{t-1} + \varepsilon_t, \text{ where} \tag{13.9}$$

$$Q_{t-1} = \begin{cases} 1 \text{ if } R_{t-1} < 0 \\ 0 \text{ else.} \end{cases}$$

This is illustrated in Figure 13.4.

Autoregressions have also been used to study the predictability of long-run returns. See Figure 13.5 for an illustration.

13.3 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.

	(1)
lag 1	−0.04
	(−2.06)
lag 2	−0.04
	(−1.77)
lag 3	−0.01
	(−0.79)
lag 4	−0.03
	(−1.69)
lag 5	−0.03
	(−1.10)
c	0.03
	(2.91)
R^2	0.00
All slopes	0.00
obs	9705.00

Table 13.1: AR(5) of daily S&P returns 1979:01-2019:12. Numbers in parentheses are t-stats, based on Newey-West with 3 lags. All slopes is the p-value for all slope coefficients being zero.

13.3.1 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 returns of large size stocks forecast returns of small size stocks. See Figure 13.6 for an illustration.

13.3.2 Earnings-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 earnings-price (or dividend-price) ratio (here in logs)

$$z^e_{s,t} = \alpha + \beta_q \ln(E_{t-s}/P_{t-s}) + \varepsilon_t, \tag{13.10}$$

where $z^e_{s,t}$ is the s-period excess log return over $t-s$ to t.

See Figure 13.7 for an illustration.

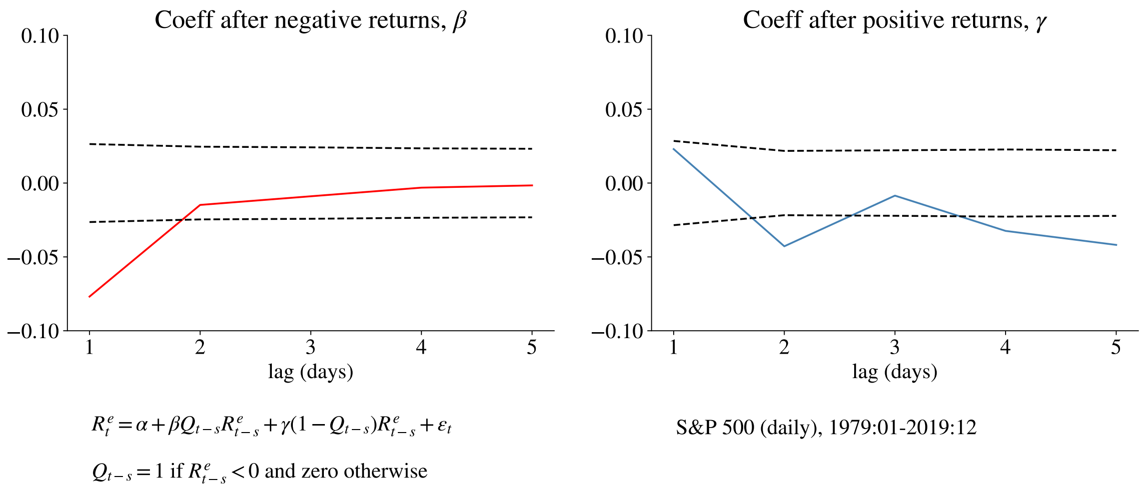


Figure 13.4: Predictability of US stock returns, results from a regression with interactive dummies

13.4 Out-of-Sample Forecasting Performance

13.4.1 In-Sample versus Out-of-Sample Forecasting

In-sample evidence on predictability can potentially be misleading because of *(a)* in-sample overfitting; and/or *(b)* structural breaks.

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 benchmark model (for instance, using the historical average as the predictor). Notice that this benchmark model is also estimated on data up to an including $t-1$, so it changes over time. See Figure 13.8.

To formalise the comparison, study the RMSE and the "out-of-sample R^2"

$$R^2_{OS} = 1 - \sum\nolimits_{t=s}^{T} (R_t - \hat{R}_t)^2 / \sum\nolimits_{t=s}^{T} (y_t - \tilde{R}_t)^2, \tag{13.11}$$

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 $t-1$).

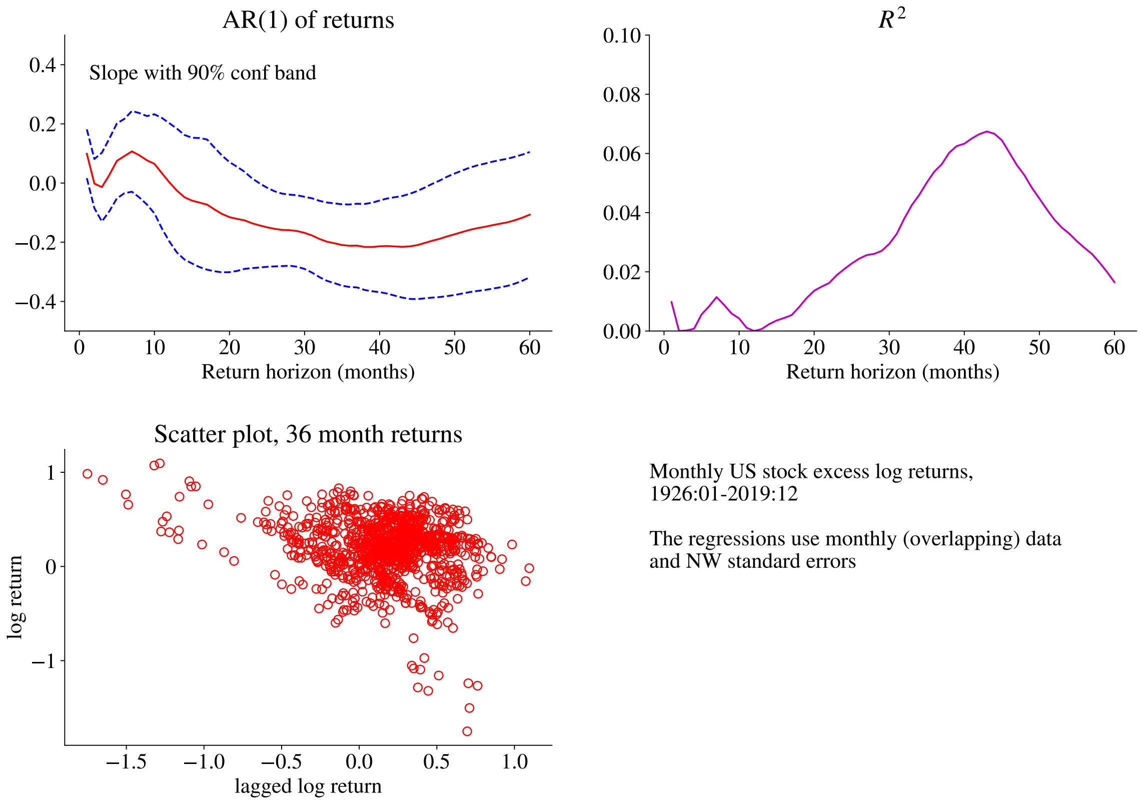


Figure 13.5: Predictability of long-run US stock returns

Example 13.3 *(R^2_{OS})*

$$R^2_{OS} = 1 - \frac{0.4}{0.5} = 0.2 \text{ (your model is better)}$$
$$R^2_{OS} = 1 - \frac{0.5}{0.4} = -0.25 \text{ (your model is worse)}$$

Goyal and Welch (2008) find that the evidence of predictability of equity returns disappears when out-of-sample forecasts are considered.

See Figures 13.9–13.10 for illustrations.

13.4.2 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 be based on the variables that are supposed to forecast returns.

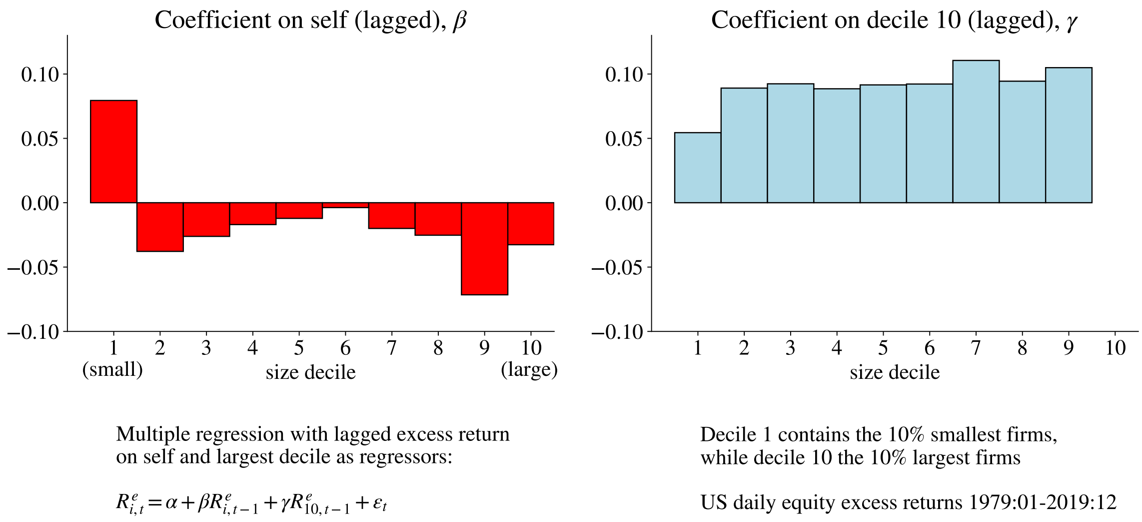


Figure 13.6: Coefficients from multiple prediction regressions

A common way to measure the performance of a portfolio is its *alpha* from a regression on the market excess return. Neutral performance requires $\alpha = 0$, which can be tested with a t test.

See Figure 13.11 for an empirical example. (In this example the alphas are almost the same as the excess return since a long-short equity portfolio has a beta close to zero.)

13.4.3 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 1990's); 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. In practice, technical analysis amounts to analysing different transformations (for instance, a moving average) of prices—and to spot 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

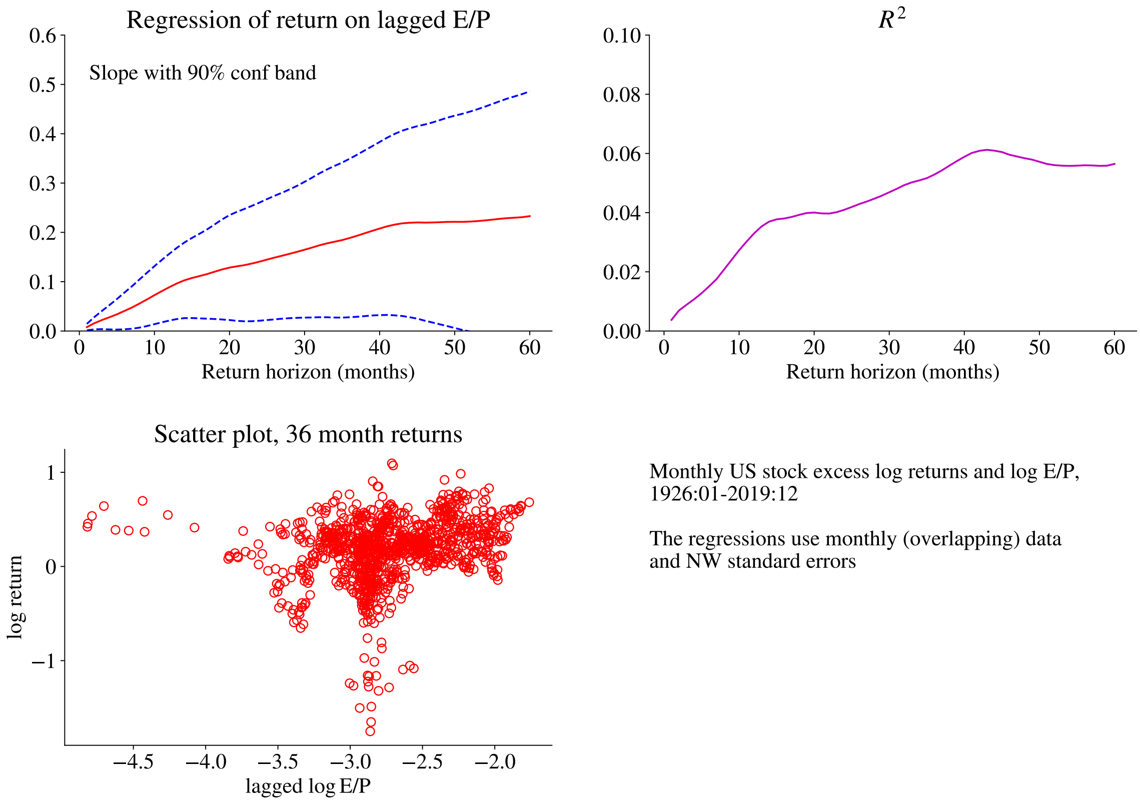


Figure 13.7: Predictability of long-run US stock returns

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 this 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

$$\begin{bmatrix} \text{buy in } t \text{ if} & MA_{t-1}(S) > MA_{t-1}(L)(1+b) \\ \text{sell in } t \text{ if} & MA_{t-1}(S) < MA_{t-1}(L)(1-b) \\ \text{no change} & \text{otherwise} \end{bmatrix}, \text{ where} \tag{13.12}$$

$$MA_{t-1}(S) = (p_{t-1} + \ldots + p_{t-S})/S.$$

The difference between the two moving averages is called an *oscillator*

$$\text{oscillator}_t = MA_t(S) - MA_t(L), \tag{13.13}$$

[1]In physics, momentum equals the mass times speed.

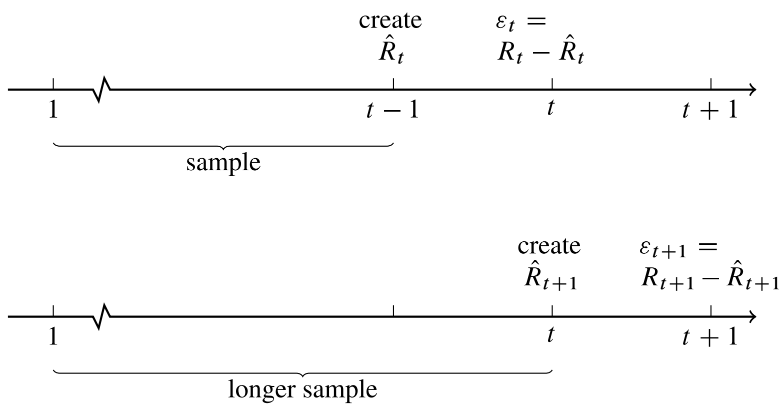


Figure 13.8: Out-of-sample forecasting

(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). A version of the moving average oscillator is the *relative strength index*[2], 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 a moving average of the oscillator (also called a signal line).

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

[2] 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).

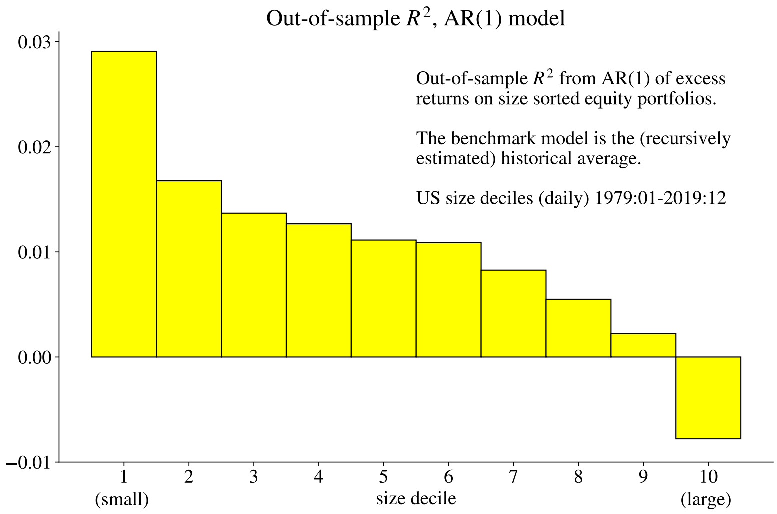


Figure 13.9: Short-run predictability of US stock returns, out-of-sample.

window of length L. With a bandwidth b (perhaps 0.01), the rule for period t could be

$$\begin{bmatrix} \text{buy in } t \text{ if} & P_t > M_{t-1}(1+b) \\ \text{sell in } t \text{ if} & P_t < m_{t-1}(1-b) \\ \text{no change} & \text{otherwise} \end{bmatrix}, \text{ where} \tag{13.14}$$

$$M_{t-1} = \max(p_{t-1}, \ldots, p_{t-S})$$
$$m_{t-1} = \min(p_{t-1}, \ldots, p_{t-S}).$$

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 13.12 and Table 13.2.

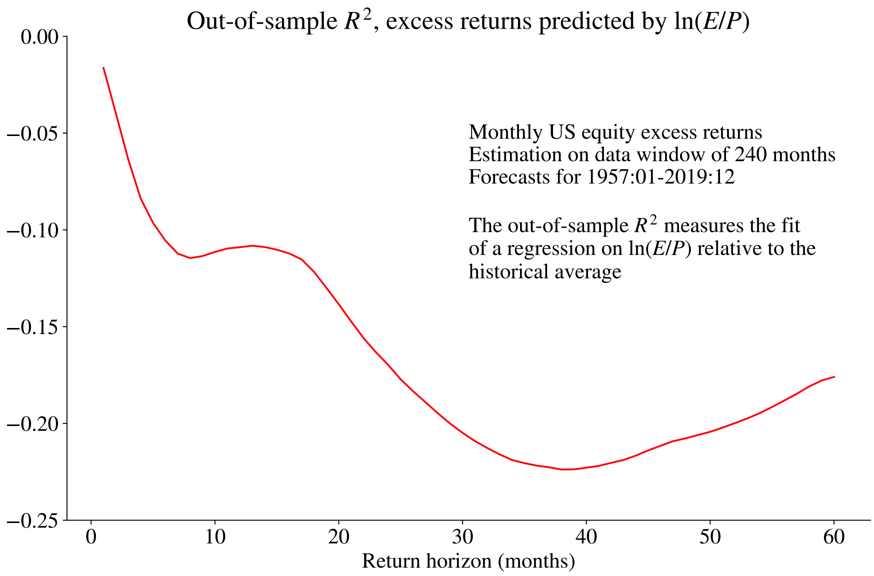


Figure 13.10: Predictability of long-run US stock returns, out-of-sample

13.5 Security Analysts

Reference: Makridakis, Wheelwright, and Hyndman (1998) 10 and Elton, Gruber, Brown, and Goetzmann (2014) 27

13.5.1 Evidence on Analysts' Performance

Makridakis, Wheelwright, and Hyndman (1998) show that there is little evidence that the average stock analyst beats (on average) the market (or 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.

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. (The number of litigations in the US after the technology boom/bust should serve as a strong reminder of this.)

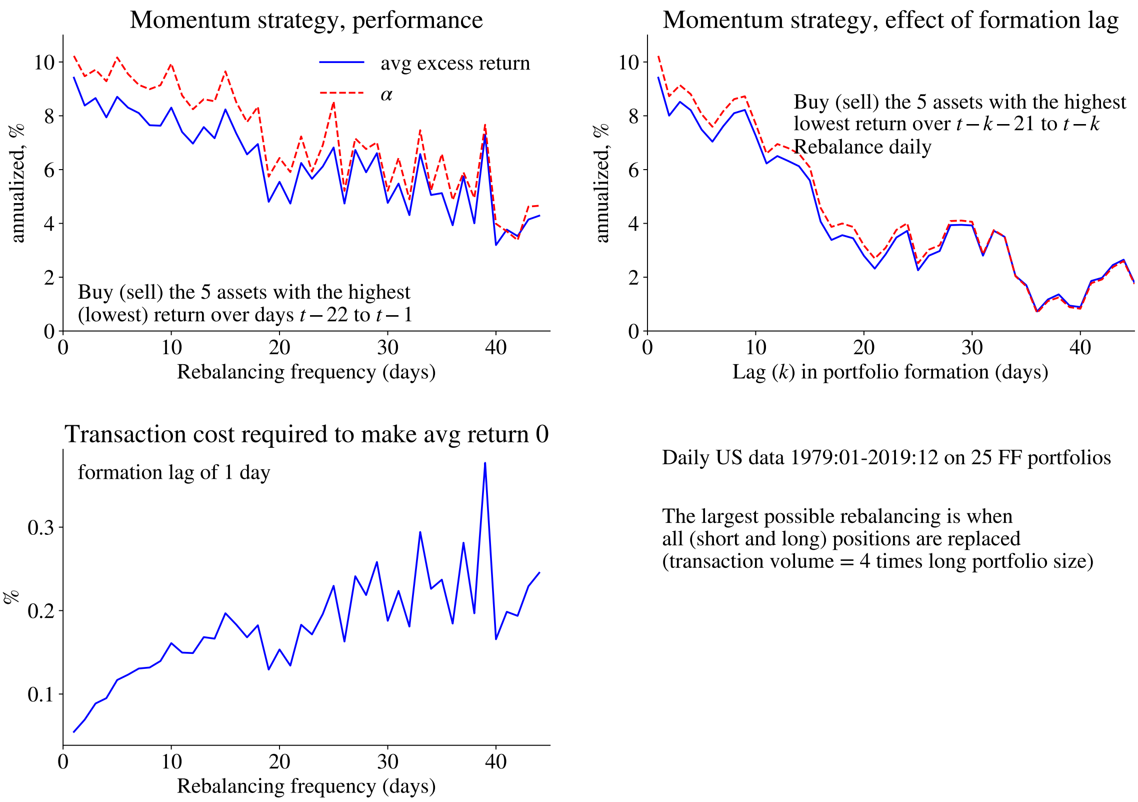


Figure 13.11: Predictability of US stock returns, momentum strategy

13.5.2 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

$$\text{Actual earnings change} = \alpha + \beta(\text{forecasted earnings change}) + \text{ residual},$$

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 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.

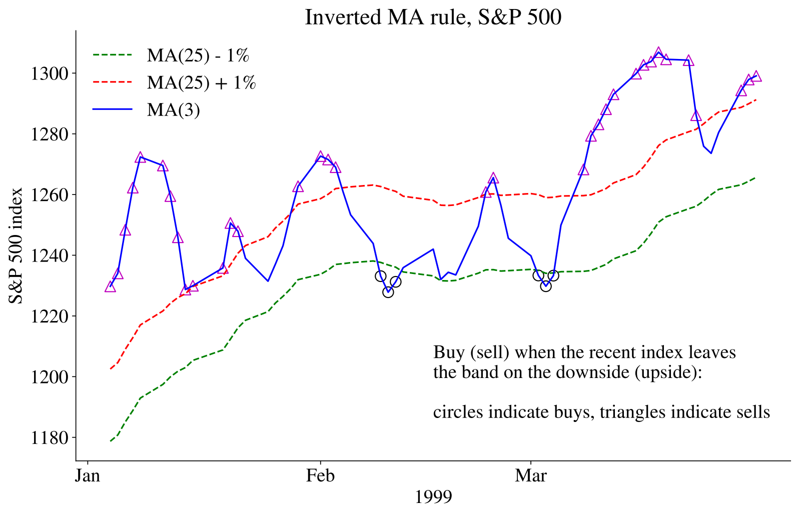


Figure 13.12: Example of a trading rule, illustration over short subsample

13.5.3 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 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."

13.5.4 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) shows that the economic experts that participate in

	Mean	Std
All days	6.5	17.3
After buy signal	17.8	26.1
After neutral signal	4.6	14.3
After sell signal	2.1	13.3
Strategy	9.6	26.1
Transaction cost	0.1	

Table 13.2: Excess returns (annualized, in %) from technical trading rule (Inverted MA rule). Daily S&P 500 data 1990:01-2019:12. The trading strategy involves (a) on everay day: hold one unit of the index and short the riskfree; (b) on days after a buy signal: double the position in (a); (c) on days after a sell signal: short sell the position in (a), effectively having a zero investment. The transaction costs shows the cost (in %) of trade that the strategy can pay and still perform as well as the static holding of (a).

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).

13.5.5 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.

13.5.6 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.

13.5.7 Mutual Funds

The general evidence on mutual funds is that they, on average, have zero alphas (or worse, after fees), and that there is little persistence in overperformance, at least among good funds(possible exceptions: hedge funds and private equity funds), while bad funds can stay bad for a long while.

13.6 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

13.6.1 Basic Structure

The idea of an event study is to study the effect of a special event by using a cross-section of such events. For instance, what is the average (across firms) effect of a negative earnings surprise on the return?

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 typically study the *abnormal return* of asset i in period t

$$u_{it} = R_{it} - R_{it}^{normal}, \tag{13.15}$$

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 13.6.2.

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 study 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

$$\bar{u}_s = \sum\nolimits_{i=1}^{n} u_{is}/n. \tag{13.16}$$

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 (13.15) 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 -20, then the 3-period (day?) car for firm i is

$$\text{car}_{i3} = u_{i,-20} + u_{i,-19} + u_{i,-18}. \tag{13.17}$$

More generally, if the event window starts at w (say, -20), then the q-period car for firm i is

$$\text{car}_{iq} = \textstyle\sum_{\tau=w}^{w+q-1} u_{i,\tau}. \tag{13.18}$$

The cross sectional average of the q-period car is

$$\overline{\text{car}}_q = \textstyle\sum_{i=1}^{n} \text{car}_{iq}/n. \tag{13.19}$$

See Figure 13.13 for an empirical example.

Example 13.4 *(Abnormal returns for ± 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

13.6.2 Models of Normal Returns

This section summarizes the most common ways of calculating the normal return in (13.15). The parameters in these models are typically estimated on a recent sample, the *estimation window*, which ends before the event window. See Figure 13.14 for an illustration. In this way, the estimated behaviour of the normal return should be unaffected by

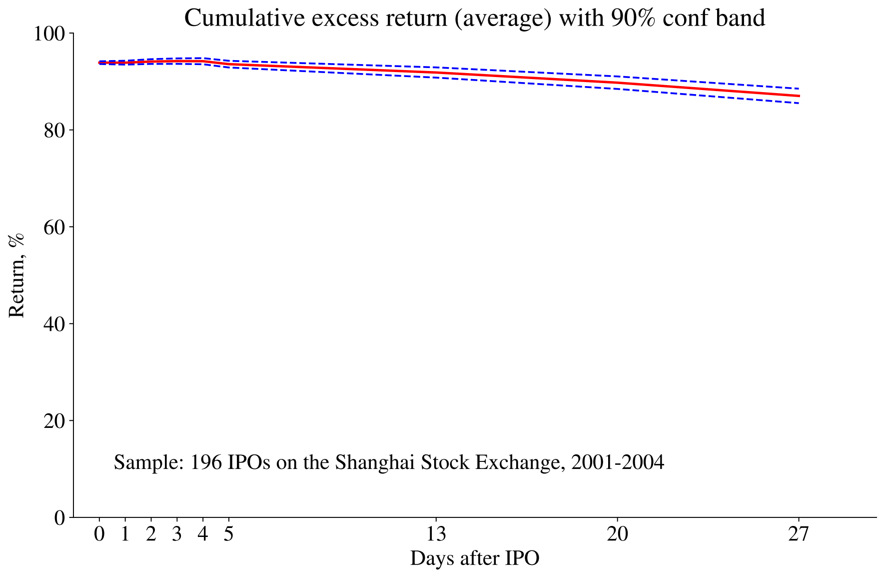


Figure 13.13: Event study of IPOs in Shanghai 2001–2004. (Data from Nou Lai.)

the event. It is almost always assumed that the event is exogenous in the sense that it is not due to the movements of 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

$$R_{it} = \mu_i + \varepsilon_{it} \text{ with} \tag{13.20}$$
$$\mathrm{E}\,\varepsilon_{it} = 0 \text{ and } \mathrm{Cov}(\varepsilon_{it}, \varepsilon_{i,t-s}) = 0.$$

This mean is estimated by the sample average (during the estimation window). The normal return in (13.15) is then the estimated mean, $\hat{\mu}_i$, so the abnormal return (in the estimation window) is the fitted residual. During the event window, we calculate the abnormal return as

$$u_{it} = R_{it} - \hat{\mu}_i. \tag{13.21}$$

The standard error of this is estimated by the standard error of the fitted residual in the estimation window.

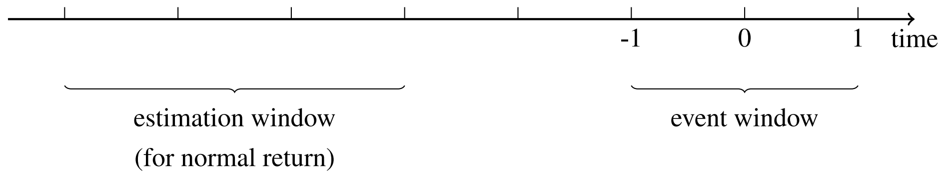


Figure 13.14: Event and estimation windows

The *market model* is a linear regression of the return of asset i on the market return

$$R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it} \text{ with} \quad (13.22)$$
$$\mathrm{E}\,\varepsilon_{it} = 0 \text{ and } \mathrm{Cov}(\varepsilon_{it}, R_{mt}) = 0.$$

Notice that we typically do not impose the CAPM restrictions on the intercept in (13.22). The normal return in (13.15) is then calculated by combining the regression coefficients with the actual market return as $\hat{\alpha}_i + \hat{\beta}_i R_{mt}$, so that the abnormal return in the estimation window is the fitted residual. For the event window we calculate the abnormal return as

$$u_{it} = R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{mt}. \quad (13.23)$$

The standard error of this is estimated by the standard error of the fitted residual 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

$$u_{it} = R_{it} - R_{mt} \quad (13.24)$$

and the standard error of it is estimated by $\mathrm{Std}(R_{it} - R_{mt})$ in the estimation window (if available).

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*. For the event window we calculate the abnormal return as

$$u_{it} = R_{it} - R_{pt}, \tag{13.25}$$

where R_{pt} is the return of the matching portfolio. The standard error of it is estimated by the standard deviation of $R_{it} - R_{pt}$ 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.

13.6.3 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 sense if the events are not overlapping in time, so that the event of assets i and j happen at different (calendar) times. If the events are overlapping, then another approach is needed.

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 (13.16), is then

$$\text{Var}(\bar{u}_s) = \textstyle\sum_{i=1}^{n} \sigma_i^2 / n^2, \tag{13.26}$$

since all covariances are assumed to be zero. In a large sample, we can therefore use a t-test since

$$\bar{u}_s / \operatorname{Std}(\bar{u}_s) \to^d N(0, 1). \tag{13.27}$$

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 $\text{car}_{i,q}$

$$\text{Var}(\text{car}_{iq}) = q\sigma_i^2. \tag{13.28}$$

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 (13.26))

$$\text{Var}(\overline{\text{car}}_q) = q \textstyle\sum_{i=1}^{n} \sigma_i^2 / n^2, \tag{13.29}$$

if the abnormal returns are uncorrelated across time and assets.

Example 13.5 *(Variances of abnormal returns) If the standard deviations of the daily abnormal returns of the two firms in Example 13.4 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 13.6 *(Tests of abnormal returns) By dividing the numbers in Example 13.4 by the square root of the numbers in Example 13.5 (that is, the standard deviations) we get the test statistic 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

Chapter 14

Dynamic Portfolio Choice

Reference: Campbell and Viceira (2002) and Merton (1973)

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

14.1 Optimal Portfolio Choice: Logarithmic Utility

Reference: Campbell and Viceira (2002)

Remark 14.1 *(Buy and hold portfolio*) A buy and hold portfolio is bought and then no further transactions are made. This means that the portfolio weights change over time. For instance, suppose* $P_{i,t} = 2$ *and* $P_{j,t} = 4$ *and we buy* 10 *of asset* i *and* 5 *of asset* j*, then the portfolio weights are* $1/2$ *for each asset (we have invested 20 into each asset). If* $P_{i,t+1} = 3$ *and* $P_{j,t+1} = 4$*, then the value in* $t+1$ *of our position in asset* i *is* $10 \times 3 = 30$ *and* $5 \times 4 = 20$ *in asset* j*. Clearly, the portfolio weights in* $t + 1$ *are* $30/50 = 0.6$ *and* $20/50 = 0.4$*, respectively.*

Remark 14.2 *(Fixed-weight portfolio*) A fixed-weight portfolio is rebalanced to keep the portfolio weights unchanged. If the prices are as in the previous remark, then we should rebalance in* $t + 1$ *so as to (again) hold equal amounts in both assets. Since the portfolio is worth* 50*, this means holding* $25/3 = 8.33$ *units of asset* i *(assuming we can buy/sell fractional amounts) and* $25/4 = 6.25$ *units of asset* j*.*

14.1.1 The Optimization Problem 1

Let the objective in period t be to maximize the expected log wealth in some future period

$$\max \mathrm{E}_t \ln W_{t+q} = \max(\ln W_t + \mathrm{E}_t\, r_{p,t+1} + \mathrm{E}_t\, r_{p,t+2} + \ldots + \mathrm{E}_t\, r_{p,t+q}), \tag{14.1}$$

where r_{pt} is the log portfolio return, $r_t = \ln(1 + R_{pt})$ and where R_{pt} is a net return. Assume the investor can rebalance the portfolio weights every period.

Since the returns in the different periods enter separably in the utility function, the best an investor can do in period t is to choose a portfolio that solves

$$\max_v \mathrm{E}_t\, r_{p,t+1}. \tag{14.2}$$

That is, to choose the one-period growth-optimal portfolio (the beliefs about $t + 2$ do not matter). This is called a *myopic* portfolio choice.

Another implication is that the investment horizon (q) does not matter: short-run and long-run investors choose the same portfolio. However, the portfolio choice may change over time (t), if the distribution of the returns does, that is, when returns are *not iid*.

14.1.2 Approximating the Log Portfolio Return

In dynamic portfolio choice models it is often 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. 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.

Remark 14.3 *("Logarithmic portfolio returns) With weights v and $1 - v$ on two assets, the gross portfolio return is $1 + R_p = ve^{r_1} + (1 - v)e^{r_2}$, where r_i is the log return on asset i. Taking logs gives $r_p = \ln[ve^{r_1} + (1 - v)e^{r_2}]$, which is clearly non-linear.*

Campbell and Viceira (2002) approximate the log portfolio return by

$$r_{pt} \approx r_{ft} + v(r_t - r_{ft}) + v\sigma^2/2 - v^2\sigma^2/2, \tag{14.3}$$

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 of the approximation is

$$r_{pt} \approx r_{ft} + v'(r_t - r_{ft}) + v' \operatorname{diag}(\Sigma)/2 - v'\Sigma v/2, \tag{14.4}$$

where Σ is the $n \times n$ covariance matrix of r_t and $\operatorname{diag}(\Sigma)$ is the $n \times 1$ vector of the variances (that is, the diagonal elements of Σ). The portfolio weights, variances and covariances could be time-varying.

Proof. (of (14.3)*) 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 $f(x) = \ln\{1 + v[\exp(x) - 1]\}$, where $x = r_1 - r_f$, has the following derivatives (evaluated at $x = 0$): $df(x)/dx = v$ and $d^2f(x)/dx^2 = v(1-v)$, 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 (14.3). (The proof of (14.4) is just a multivariate extension of this.) ■

14.1.3 The Optimization Problem 2

The objective is to maximize the (conditional) expected value of the portfolio return as in (14.2), $\max \mathrm{E}_t\, r_{p,t+1}$. When there is one risky asset and a riskfree asset, then the portfolio return is given by the approximation (14.3). To simplify the notation a bit, let μ^e be the conditional expected excess log return of the risky asset (that is, $\mathrm{E}_t(r_{t+1} - r_{f,t+1})$) and let σ^2 be its conditional variance (that is, $\mathrm{Var}_t(r_{t+1})$). 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 to maximize $\max \mathrm{E}_t\, r_{p,t+1}$, which is (approximately, according to (14.3)) the same as

$$\max\nolimits_{v_t} r_{f,t+1} + v\mu^e + v\sigma^2/2 - v_t^2\sigma^2/2. \tag{14.5}$$

The first order condition is

$$0 = \mu^e + \sigma^2/2 - v\sigma^2, \text{ so}$$
$$v = \frac{\mu^e + \sigma^2/2}{\sigma^2}, \tag{14.6}$$

and where $1 - v$ is invested in the riskfree asset. Clearly, the portfolio weight v changes over time—if the expected excess return and/or the volatility does, that is, when returns are not iid. We could think of the portfolio as a *managed portfolio*.

Example 14.4 *(One risky asset) Suppose there is one risky asset with $\sigma^2 = 0.01$, and the expected excess returns are different the two "states" (A and B). Then (14.6) gives.*

State	μ^e	v
A	$1/100$	1.5
B	$0.5/100$	1.0

The weight on the riskfree asset is clearly -0.5 in state A and 0 in state B.

Remark 14.5 *(*Comparison with standard MV results) The optimal portfolio weight (14.6) is very similar to what we get from maximizing $\mathrm{E}(vR^e + R_f) - \mathrm{Var}(vR^e + R_f)/2$, which gives $v = \mathrm{E}\, R^e / \mathrm{Var}(R^e)$. The $\sigma^2_{t+1}/2$ in the numerator of (14.6) is easily understood once we remember that if $r \sim N(\mu, \sigma^2)$, then $\mathrm{E}(1 + R) = \exp\left(\mu + \sigma^2/2\right)$.*

With many risky assets, the optimization problem is to maximize the expected value of (14.4). The optimal $n \times 1$ vector of portfolio weights is then

$$v = \Sigma^{-1}(\mu^e + \mathrm{diag}(\Sigma)/2), \tag{14.7}$$

where Σ is the conditional variance-covariance matrix ($\mathrm{Cov}_t(r_{t+1})$) and $\mathrm{diag}(\Sigma)$ the $n \times 1$ vector of conditional variances (the diagonal of Σ). The weight on the riskfree asset is the remainder ($1 - \mathbf{1}'v$, where $\mathbf{1}$ is a vector of ones).

Proof. (*of (14.7)) From (14.4) we have that the objective function can be written

$$\mathrm{E}\, r_p + (1 - \gamma)\, \mathrm{Var}(r_p)/2 \approx r_f + v'\mu^e + v'\, \mathrm{diag}(\Sigma)/2 - v'\Sigma v/2,$$

so the first order conditions are

$$\mu^e + \mathrm{diag}(\Sigma)/2 - \Sigma v = \mathbf{0}_{n\times 1}.$$

Solve for v. ■

Example 14.6 *(Three risky assets) Suppose we have three assets with the variance-covariance matrix (which is the same in both states)*

$$\Sigma = \begin{bmatrix} 166 & 34 & 58 \\ 34 & 64 & 4 \\ 58 & 4 & 100 \end{bmatrix} /10000,$$

and the means (in state A and B, respectively)

$$\mu_A^e = \begin{bmatrix} 2 \\ 1 \\ 0.5 \end{bmatrix} /100 \text{ and } \mu_B^e = \begin{bmatrix} 2 \\ 0 \\ 0.5 \end{bmatrix} /100,$$

In this case, the portfolio weights in the two states are

$$v_A \approx \begin{bmatrix} 1.38 \\ 1.32 \\ 0.14 \end{bmatrix} \text{ and } v_B \approx \begin{bmatrix} 1.81 \\ -0.46 \\ -0.03 \end{bmatrix}.$$

Figure 14.1 illustrates mean returns and standard deviations, estimated by exponentially weighted moving averages (as by RiskMetrics). Figure 14.2 shows how the optimal portfolio weights change. It is clear that the portfolio weights can be fairly extreme and also change a lot—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.

14.2 Optimal Portfolio Choice: CRRA Utility

The previous section has shown that logarithmic utility leads to myopic behaviour where the optimal portfolio depends only on beliefs about the next-period return. This clearly simplifies the choice, but it is unclear if logarithmic utility is a good representation of preferences. We therefore extend the analysis to the general constant relative risk aversion (CRRA) case.

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

$$\mathrm{E}\, r_p + (1 - \gamma)\,\mathrm{Var}(r_p)/2, \tag{14.8}$$

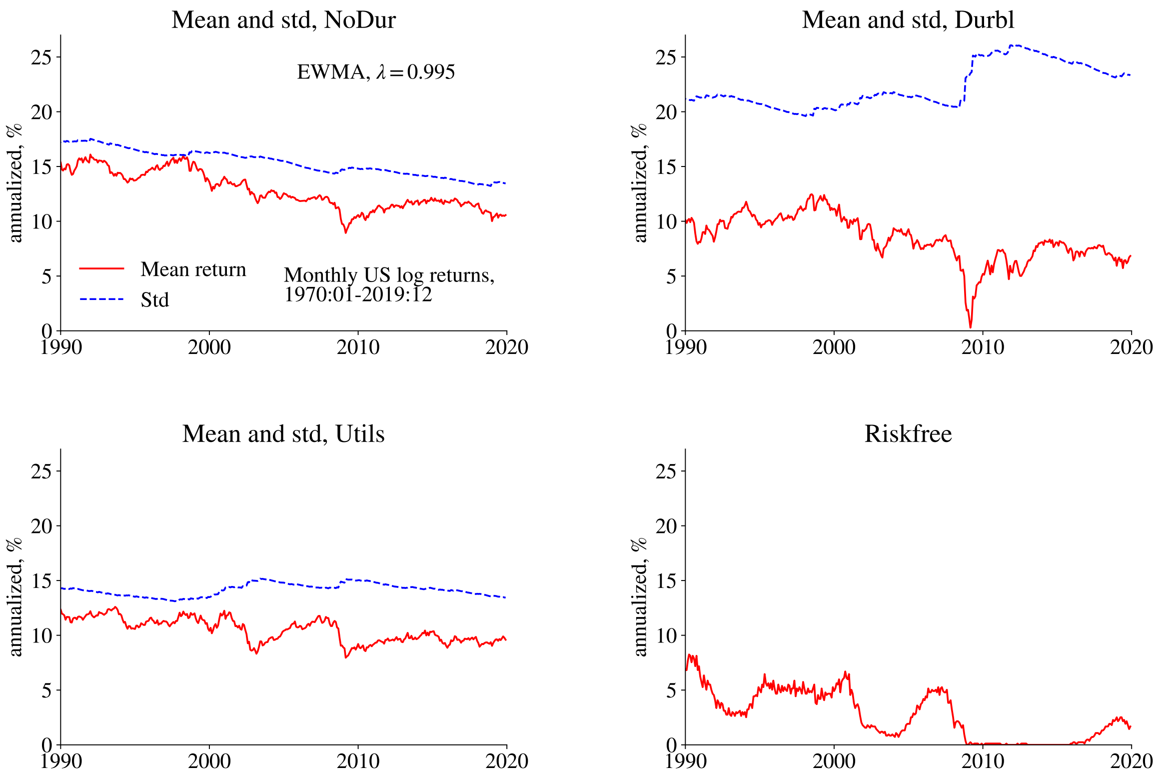


Figure 14.1: Dynamically updated estimates, 5 U.S. industries

where r_p is the log return of the portfolio (strategy) over the investment horizon. (See lecture notes in utility theory for a proof.)

Maximizing (14.8) gives the following vector of optimal portfolio weights

$$v = \Sigma^{-1}(\mu^e + \text{diag}(\Sigma)/2)/\gamma. \tag{14.9}$$

With only one risky asset, Σ is just a scalar (a variance) and the same as $\text{diag}(\Sigma)$, so the expression simplifies.

Proof. (*of (14.9)) From (14.4) we have that the objective function (14.8) can be written

$$\text{E}\, r_p + (1-\gamma)\,\text{Var}(r_p)/2 \approx r_f + v'\mu^e + v'\,\text{diag}(\Sigma)/2 - v'\Sigma v/2 + (1-\gamma)v'\Sigma v/2,$$

so the first order conditions are

$$\mu^e + \text{diag}(\Sigma)/2 - \gamma\Sigma v = \mathbf{0}_{n\times 1}.$$

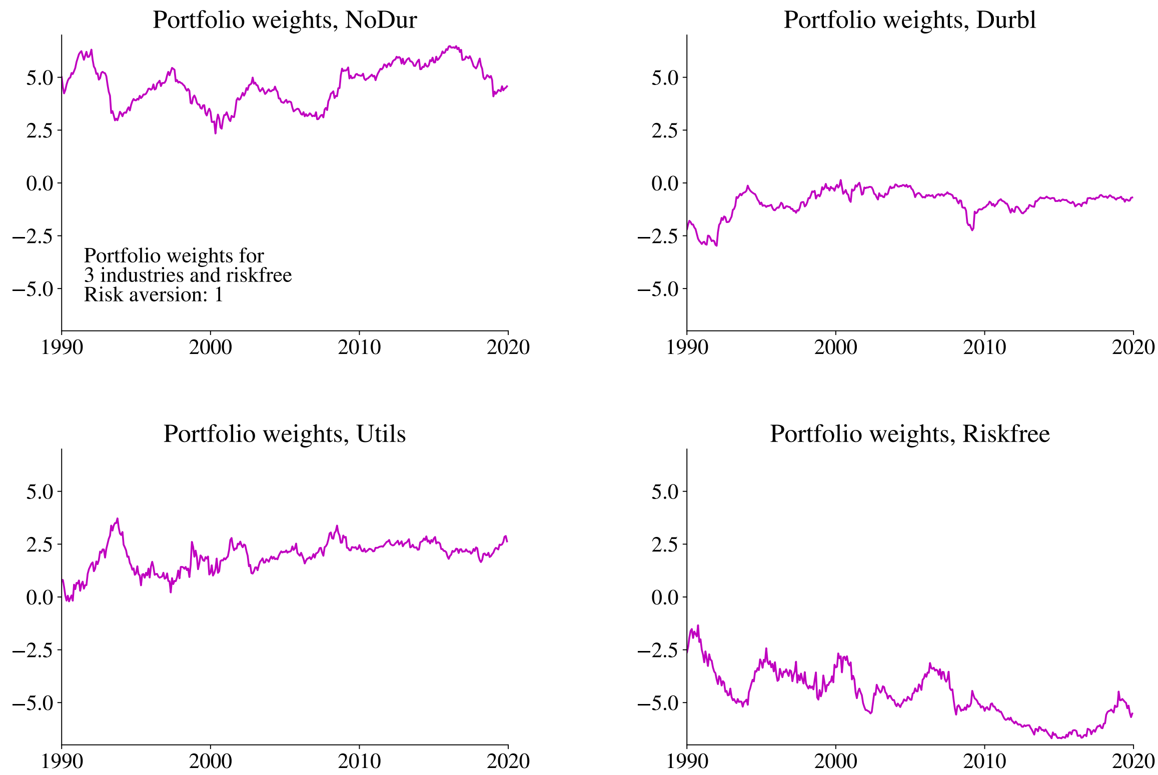


Figure 14.2: Dynamically updated portfolio weights, T-bill and 5 U.S. industries

Solve for v. ■

Example 14.7 *(One risky asset) Using the same figures as in Example 14.4 and* $\gamma = 3$ *gives* $v_A = 0.5$ *and* $v_B = 0.33$.

Example 14.8 *(Three risky assets) Using the same figures as in Example 14.6 and* $\gamma = 3$ *gives*

$$v_A \approx \begin{bmatrix} 0.46 \\ 0.44 \\ 0.05 \end{bmatrix} \text{ and } v_B \approx \begin{bmatrix} 0.60 \\ -0.15 \\ -0.01 \end{bmatrix}.$$

14.3 Intertemporal Hedging (CRRA Utility and non-iid Returns)

14.3.1 Basic Setup

However, the *combination* of a utility function with a CRRA different from unity (the log utilkity case) and non-iid returns complicates the portfolio choice. In particular, the portfolio choice in period t might 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, we l illustrate it by using a simple model (see Campbell and Viceira (1999) for sophisticated models).

Suppose the vector (n assets) of excess returns follow

$$r^e_{t+1} = a + z_t + u_{t+1}, \tag{14.10}$$

where a is an n-vector of long-run averages, u_{t+1} is an n-vector of $N(\mathbf{0}, \Sigma_u)$ shocks and where the n-vector of signals z_t (information variables) follows a VAR(1)

$$z_{t+1} = \phi z_t + \eta_{t+1} \tag{14.11}$$

where η_{t+1} is $N(\mathbf{0}, \Sigma_\eta)$. We denote the covariance matrix of u_{t+1} and η_{t+1} by $\Sigma_{u\eta}$. This model has time-varying expected returns and allows the *return shocks* to be correlated with *expected future returns*.

When an element of $\Sigma_{u\eta} < 0$, then a positive return shock ($u_{t+1} > 0$) is typically accompanied by a negative shock (η_{t+1}) to the signal, which means that future expected returns are negatively affected ($\mathrm{E}_{t+1}\, r^e_{t+2}$ decreases). This is the key to get intertemporal hedging.

In contrast, when $z_t = 0$ in all periods (because there are no η_t shocks, meaning $\Sigma_\eta = \mathbf{0}$), then the excess returns are iid.

Remark 14.9 *(Expected returns and innovations in returns) From (14.10)–(14.11), we have*

$$\mathrm{E}_t\, r_{t+1} = a + z_t, \text{ so } r_{t+1} - \mathrm{E}_t\, r_{t+1} = u_{t+1}$$
$$\mathrm{E}_t\, r_{t+2} = a + \phi z_t, \text{ so } r_{t+2} - \mathrm{E}_t\, r_{t+2} = \eta_{t+1} + u_{2+1}$$

This means that $\mathrm{Cov}_t(r_{t+1}) = \Sigma_u$ *and* $\mathrm{Cov}_t(r_{t+2}) = \Sigma_\eta + \Sigma_u + 2\Sigma_{\eta u}$.

14.3.2 Myopic Portfolio Choice

A *myopic investor* (one-period investor) maximizes (14.8), but where expectation and the variance are for a 1-period return ($\mathrm{E}_t\, r_{pt+1}$ and $\mathrm{Var}_t(r_{pt+1})$ respectively), that is,

$$\max_v \mathrm{E}_t\, r_{pt+1} + (1-\gamma)\,\mathrm{Var}_t(r_{pt+1})/2. \tag{14.12}$$

Using the results from Remark 14.9 in (14.9) gives the optimal portfolio

$$v = \Sigma_u^{-1}(a + z_t + \text{diag}(\Sigma_u)/2)/\gamma. \tag{14.13}$$

With $\gamma = 1$ we get the result for log utility (the same as in (14.6), but with slightly different notation). With a higher risk aversion, the weights on the risky asset are closer to zero. As before, high expected excess returns $(a + z_t)$ typically give high weights on the risky assets.

See Figures 14.3 and 14.4 for illustrations.

Example 14.10 *(One risky asset) With one risky asset and* $(\gamma, a, \Sigma_u, \Sigma_\eta) = (3, 0.75/100, 100/10000, 75/10000)$, *we get* $v = 0.42$ *when* $z_t = 0$ *(the average value).*

Example 14.11 *(Several risky assets) Let there be three risky assets with*

$$\Sigma_u = \begin{bmatrix} 166 & 34 & 58 \\ 34 & 64 & 4 \\ 58 & 4 & 100 \end{bmatrix} /10000,\ \Sigma_\eta = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 0 \end{bmatrix} /10000$$

and the average returns

$$a = \begin{bmatrix} 2 \\ 0.5 \\ 0.5 \end{bmatrix} /100.$$

The myopic portfolio weights are

$$v \approx \begin{bmatrix} 0.53 \\ 0.14 \\ 0.02 \end{bmatrix}.$$

14.3.3 A Two-Period Investor (No Rebalancing)

In period t, a two-period investor also maximizes (14.8), but the expectation and variance are for a 2-period return ($\text{E}_t(r_{pt+1} + r_{pt+2})$ and $\text{Var}_t(r_{pt+1} + r_{pt+2})$ respectively), that is,

$$\max_v \text{E}_t(r_{pt+1} + r_{pt+2}) + (1 - \gamma)\,\text{Var}_t(r_{pt+1} + r_{pt+2})/2. \tag{14.14}$$

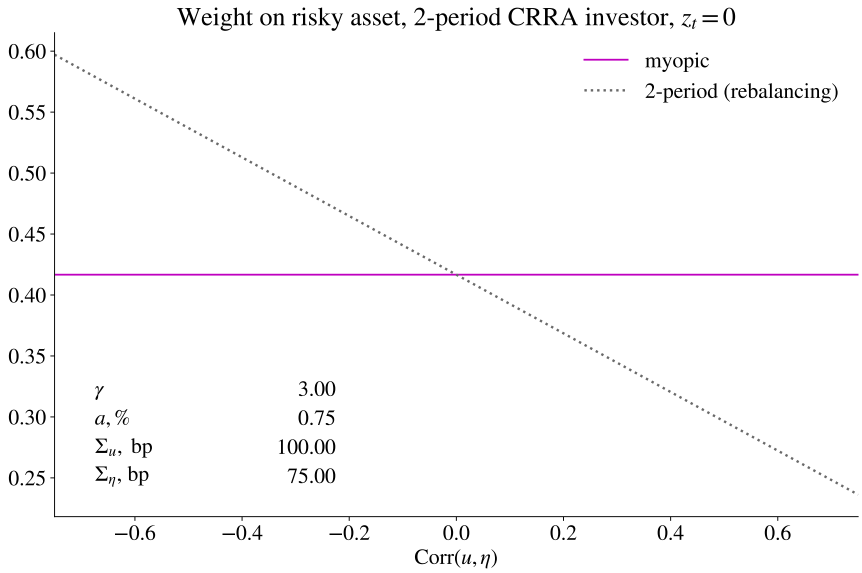


Figure 14.3: Weight on risky asset, two-period investor with CRRA utility and the possibility to rebalance

The optimal portfolio choice is

$$v = \Sigma^{-1}[2a + (I + \phi)z + \text{diag}(\dot{\Sigma})/2]/\gamma, \text{ where} \tag{14.15}$$
$$\Sigma = 2\Sigma_u + \Sigma_\eta + 2\Sigma_{u\eta}.$$

Compared to the 1-period (or myopic) investor, this 2-period investor considers the risk differently. In particular, the η_{t+1} shocks cause more uncertainty (about r_{t+2}) which is captured by the diagonal elements in Σ_η. This is the same as saying that long-run returns are riskier than short-run returns. On the other hand, if u_{t+1} and η_{t+1} are negative correlated (negative elements in $\Sigma_{u\eta}$), then the long-run risk is dampended by the mean reversion: a shock to r_{t+1} will be partly offset by a movement in r_{t+2} of the opposite sign.

Proof. (of (14.15)) Remark 14.9 gives $\text{E}_t(r_{pt+1} + r_{pt+2}) = 2a + (I + \phi)z_t$ and $\text{Cov}_t(r_{1+2} + r_{t+2}) = \text{Cov}(u_{t+1} + \eta_{t+1} + u_{2+1}) = 2\Sigma_u + \Sigma_\eta + 2\Sigma_{u\eta}$. Use in (14.9). ■

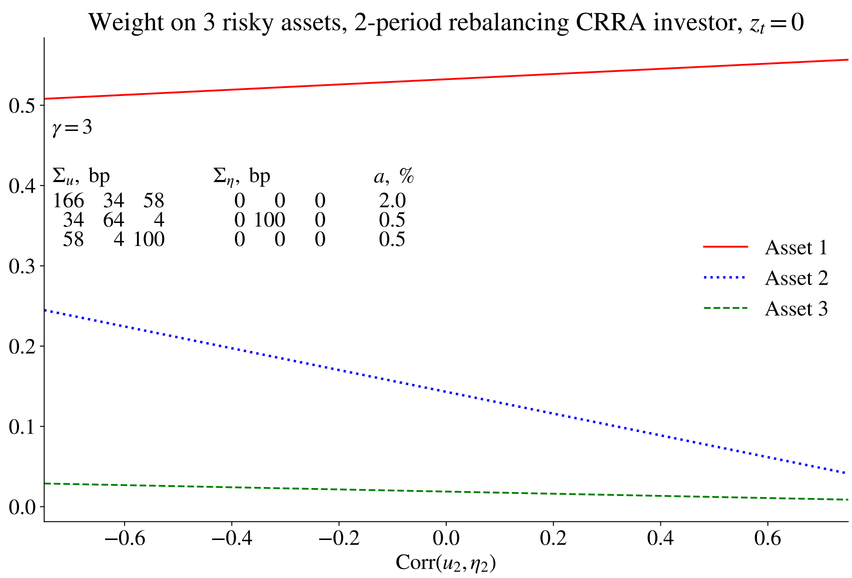


Figure 14.4: Weight on risky asset, two-period investor with CRRA utility and the possibility to rebalance

14.3.4 Two-Period Investor (with Rebalancing)

It is perhaps more reasonable to assume that the two-period investor can rebalance in each period. Rewrite the 2-period problem (14.14) as

$$\mathrm{E}_t\, r_{pt+1} + \mathrm{E}_t\, r_{pt+2} + (1-\gamma)[\mathrm{Var}_t(r_{pt+1}) + \mathrm{Var}_t(r_{pt+2}) + 2\,\mathrm{Cov}_t(r_{pt+1}, r_{p2+1})]/2. \tag{14.16}$$

When choosing the portfolio in period t, the investor cannot directly influence $\mathrm{E}_t\, r_{pt+2}$ and $\mathrm{Var}_t(r_{pt+2})$, since they depend on the portfolio chosen in $t+1$. (The no-rebalancing case can be thought of as pre-committing to a next period portfolio choice.)

In contrast, the portfolio choice in t directly affects $\mathrm{E}_t\, r_{pt+1}$,$\mathrm{Var}_t(r_{pt+1})$ and $\mathrm{Cov}_t(r_{pt+1}, r_{p2+1})$. The first two terms are just like in the myopic case, but the covariance is trickier.

In principle, the covariance term is

$$\mathrm{Cov}_t(r_{pt+1}, r_{p2+1}) = v'\,\mathrm{Cov}_t(r^e_{t+1}, r^e_{t+2} v_{t+1}), \tag{14.17}$$

where v is the portfolio choice in t (and can therefore be moved outside the covariance operator) and v_{t+1} in $t+1$. The latter must be a 1-period (myopic) choice as in (14.13), but applied to $t+1$ (substitute z_{t+1} for z_t). Clearly, those weights are not known in t. This shows that the covariance (14.17) is tricky since $r^e_{t+2} v_{t+1}$ includes products of random variables. While it can be solved, the expressions are long. We therefore approximate the covariance by

$$\mathrm{Cov}_t(r_{pt+1}, r_{p2+1}) \approx v' \Sigma_{u\eta} \, \mathrm{E}_t \, v_{t+1}, \tag{14.18}$$

which means that we treat v_{t+1} as being known in t (and equal to $\mathrm{E}_t \, v_{t+1}$), although decided in $t+1$.

Combining into a new objective function gives

$$\mathrm{E}_t \, r_{pt+1} + (1-\gamma)[\mathrm{Var}_t(r_{pt+1})/2 + v' \Sigma_{u\eta} \, \mathrm{E}_t \, v_{t+1}], \tag{14.19}$$

with the optimal portfolio choice

$$v = \Sigma_u^{-1}[a + z_t + \mathrm{diag}(\Sigma_u)/2 + (1-\gamma)\Sigma_{u\eta} \, \mathrm{E}_t \, v_{t+1}]/\gamma. \tag{14.20}$$

This equals the myopic portfolio 14.13 is two cases: (1) when $\gamma = 1$ (log utility); and when (2) $\Sigma_{u\eta} = \mathbf{0}$ (no mean reversion or autocorrelation of returns).

In other cases, a $\gamma > 1$ and negative elements in $\Sigma_{u\eta}$ will contribute to a higher portfolio weight. That is, assets that pays off in $t+1$ when the outlook for $t+2$ worsens (and vice versa) provide a diversification benefit—and this increases the demand for them. This is often called *intertemporal hedging.*

Figures 14.3 illustrates that the weight on a single risky asset decreases as $\mathrm{Corr}(u, \eta)$ increases. When the correlation is zero, then we get the same weight as in the myopic case. Similarly, Figure 14.4 illustrates how one of the risky assets (asset 2) becomes less demanded when $\mathrm{Corr}(u, \eta)$ for that asset increases. Again, it can be show that the $\mathrm{Corr}(u, \eta) = 0$ replicates the myopic portfolio choice.

Proof. (*of (14.20)) Add the $(1-\gamma)\,\mathrm{Cov}_t(r_{pt+1}, r_{p2+1})$ term to (14.8). Following the same approach as in the proof of (14.9) gives

$$r_f + v'\mu^e + v' \mathrm{diag}(\Sigma)/2 - v'\Sigma v/2 + (1-\gamma)v'\Sigma v/2 + (1-\gamma)v'\Sigma_{u\eta} \, \mathrm{E}_t \, v_{t+1}.$$

The first order conditions are

$$\mu^e + \mathrm{diag}(\Sigma)/2 - \gamma\Sigma v + (1-\gamma)\Sigma_{u\eta} \, \mathrm{E}_t \, v_{t+1} = \mathbf{0}_{n\times 1}.$$

Substitute $a + z_t$ for μ^e, and Σ_u for Σ and solve for v. ■

Example 14.12 *(One risky asset) Using the same figures as in Example 14.10 we get* $v = 0.54$ *when* $\mathrm{Corr}(u, \eta) = -0.5$ *and* $v = 0.30$ *when* $\mathrm{Corr}(u, \eta) = 0.5$.

Example 14.13 *(Several Risky assets) Using the same figures as in Example 14.11 and also two possible $\Sigma_{u\eta}$: with positive and negative correlation of u and η for asset 2 (zero for all other assets)*

$$\Sigma_{u\eta} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -40 & 0 \\ 0 & 0 & 0 \end{bmatrix} /10000 \; or \begin{bmatrix} 0 & 0 & 0 \\ 0 & 40 & 0 \\ 0 & 0 & 0 \end{bmatrix} /10000.$$

The portfolio weights in two cases are

$$v \approx \begin{bmatrix} 0.52 \\ 0.21 \\ 0.03 \end{bmatrix} \; or \; \begin{bmatrix} 0.55 \\ 0.08 \\ 0.01 \end{bmatrix}.$$

Bibliography

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

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.

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.

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

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.

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.

Brandimarte, P., 2006, *Numerical Methods in Finance and Economics*, Wiley, Hoboken, NJ.

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.

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 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.

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.

Chen, N.-F., R. Roll, and S. A. Ross, 1986, "Economic forces and the stock market," *Journal of Business*, 59, 383–403.

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.

Danthine, J.-P., and J. B. Donaldson, 2005, *Intermediate financial theory*, Elsevier Academic Press, 2nd edn.

Elton, E. J., M. J. Gruber, S. J. Brown, and W. N. Goetzmann, 2014, *Modern portfolio theory and investment analysis*, John Wiley and Sons, 9th 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.

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

Forbes, W., 2009, *Behavioural finance*, Wiley.

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.

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

He, J., and L. Ng, 1994, "Economic forces and the stock market," *Journal of Business*, 4, 599–609.

Huang, C.-F., and R. H. Litzenberger, 1988, *Foundations for financial economics*, Elsevier Science Publishing, New York.

Hull, J. C., 2009, *Options, futures, and other derivatives*, Prentice-Hall, Upper Saddle River, NJ, 7th edn.

Ibbotson, R. G., and P. D. Kaplan, 2000, "Does Asset Allocation Policy Explain 40, 90 or 100 Percent of Performance?," *Financial Analysts Journal*, 65, 26–33.

Ingersoll, J. E., 1987, *Theory of financial decision making*, Rowman and Littlefield.

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.

McDonald, R. L., 2014, *Derivatives markets*, Pearson, 3rd edn.

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

Merton, R. C., 1973, "An intertemporal capital asset pricing model," *Econometrica*, 41, 867–887.

Murphy, J. J., 1999, *Technical analysis of the financial markets*, New York Institute of Finance.

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.

Neely, C. J., 1997, "Technical analysis in the foreign exchange market: a layman's guide," *Federal Reserve Bank of St. Louis Review*.

Ross, S. A., 1976, "The arbitrage pricing theory," *Journal of Economic Theory*, 1, 341–360.

Sercu, P., 2009, *International Finance*, Princeton University Press.

Sharpe, W. F., 1992, "Asset allocation: management style and performance measurement," *Journal of Portfolio Management*, 39, 119–138.

Shefrin, H., 2005, *A behavioral approach to asset pricing*, Elsevier Academic Press, Burlington, MA.

Söderlind, P., 2010, "Predicting stock price movements: regressions versus economists," *Applied Economics Letters*, 17, 869–874.

Stefek, D., 2002, "The Barra integrated model," *Barra Research Insight*.

The Economist, 1993, "Frontiers of finance," pp. 5–20.

Verbeek, M., 2004, *A guide to modern econometrics*, Wiley, Chichester, 2nd edn.