Modern Finance Theory

Paul Söderlind[1]

26 December 2025

[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: FinAll.TeX.

These lecture notes are for a first M.A. course in finance. The goal is to present financial theory in a way so that it can be used directly in quantitative/empirical projects that require numerical estimations and computations. The approach is therefore formal, but the mathematics is relatively easy, for instance, linear algebra is used, but stochastic calculus is not. Also, in terms of scope, the focus is on classical concept, for instance, the tangency portfolio plays an important role, but pricing kernels do not. Optional (often more advanced) material is denoted by a star*.

In implementing numerical computations based on these notes, my students have typically used Julia, Matlab, Python or R. Julia notebooks with numerical examples for each chapter are found at Paul Söderlind's Github page: `https://github.com/PaulSoderlind/FinancialTheoryMSc`

When I first set up this course many years ago, I was inspired by the texts of Bodie, Danthine&Donaldson, Elton&Gruber, and Hull. Most likely that still shows.

My students at the MiQEF program at the University of St. Gallen have asked many good questions and pointed out mistakes. Also my teaching assistants did the same. Without that, these notes would have been worse.

Data Sources

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

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

Contents

Chapter 1

The Basics of Return Calculations

This chapter first defines *returns*, demonstrates how to summarise their statistical properties (descriptive statistics), and discusses how to accumulate them. It then shows how *portfolio returns* depend on the returns of the assets in the portfolio. Later sections summarise the basic statistical properties of important asset classes and some key markets and trading concepts.

The material in this chapter is presented in a dense format; it is basically a list of definitions with some short comments, and probably more useful for looking up a specific definition than for reading page by page.

1.1 Asset Returns

1.1.1 Definition of a Return

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

$$R_t = \frac{V_t - V_{t-1}}{V_{t-1}} = \frac{V_t}{V_{t-1}} - 1, \tag{1.1}$$

where V_t is the value of the asset in period t.

Remark 1.1 *(On notation) A precise notation (which time, investment horizon, asset, units. ...) can be cumbersome. When needed, we will use R_{it} (but $R_{i,t-1}$) to indicate the return of asset i in period t ($t-1$). However, when dealing with a single asset where the time dimension is important, then we just keep the t subscript. Instead, when dealing with several assets but where the time dimension is less important, then we just keep the i subscript. Sometimes we drop all subscripts. The meaning should be clear from the context.*

The gross return is

$$1 + R_t = \frac{V_t}{V_{t-1}}. \tag{1.2}$$

Example 1.2 *(Returns)*

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

Remark 1.3 *(% 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

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

so the return can be written

$$\begin{aligned} R_t &= \frac{D_t + P_t - P_{t-1}}{P_{t-1}} \\ &= \underbrace{\frac{D_t}{P_{t-1}}}_{\text{dividend yield}} + \underbrace{\frac{P_t - P_{t-1}}{P_{t-1}}}_{\text{capital gain yield}} \end{aligned} \tag{1.4}$$

Example 1.4 *(Dividend yield ad capital gain yield)*

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

The *excess return* of an asset (compared to the risk-free rate R_f) is

$$R_t^e = R_t - R_{ft}. \tag{1.5}$$

In other cases, we use something else than a risk-free return as the reference rate.

Example 1.5 *(Excess return) If* $R_t = 0.08$ *and* $R_{ft} = 0.01$*, then the excess return is* $R_t^e = 0.07$ *(7%).*

Remark 1.6 *(Approximating the risk-free return*) Suppose you have monthly equity returns and want to calculate excess returns. Do as follows. First, find a repre-*

sentative money market instrument (for instance, a T-bill or an interbank contract) with approximately one month to maturity. Second, use the interest rate on that instrument quoted a month ago. divided by 1. (This is the rate you earn/pay on a loan between a month ago and now.) 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.

1.1.2 Logarithmic Returns*

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 somewhat inconvenient when the focus is on choosing the portfolio weights: the log portfolio return is *not* a weighted average of the log returns of the assets in the portfolio. (An approximation might work, as demonstrated in the chapter on dynamic portfolio choice.)

Anyhow, a log return is defined as

$$r_t = \ln(1 + R_t), \tag{1.6}$$

which clearly equals $\ln(V_t / V_{t-1})$. To convert from log returns to net returns, use $R_t = \exp(r_t) - 1$.

The corresponding excess log return is

$$r_t^e = \ln(1 + R_t) - \ln(1 + R_{ft}). \tag{1.7}$$

Assuming we invest an equal amount in both instruments in $t - 1$ ($V_{t-1} = V_{f,t-1}$), the excess log return equals $\ln(V_t / V_{ft})$. Notice that excess log return is *not* the log of the excess return. Rather, it is the log of $(1 + R_t)/(1 + R_{ft})$. Figure 1.1 illustrates that the difference between r^e and the possible approximation $\ln(1 + R^e)$ can be substantial.

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

1.1.3 Inflation and Real Returns

In most portfolio choice models, it is the *real* return (measured in units of "goods") that matters, not the *nominal* return (measured in currency units). The reason is

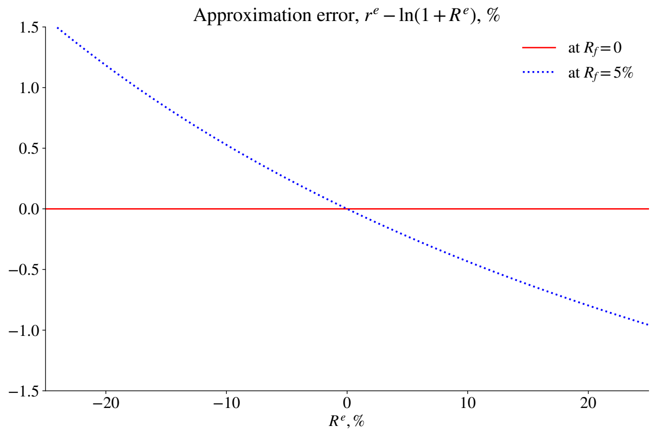


Figure 1.1: Approximation error from using $\ln(1+R^e)$ instead of r^e

straightforward: utility depends on real goods and services, not on nominal price levels.

To see the link between real and nominal returns, let Γ_t be the nominal price level (price of the consumption basket of an investor, measured in currency units). If V in (1.1) is a nominal value (measured in currency units), then the real value is V/Γ.

Example 1.8 *(Nominal and real prices) If* $V = 110$ *is the nominal value of an asset and* $\Gamma = 5$ *is the nominal price of the consumption basket, then the real value is* $V/\Gamma = 22$*. This represents the number of consumption baskets required to match the asset's value.*

The real return (corresponding to (1.1)) is

$$\tilde{R}_t = \frac{\Gamma_{t-1}}{\Gamma_t}\frac{V_t}{V_{t-1}} - 1$$
$$= \frac{1+R_t}{1+\pi_t} - 1, \tag{1.8}$$

where $\pi_t = \Gamma_t/\Gamma_{t-1} - 1$ is the inflation rate. We get a similar expression for the risk-free rate, so the excess real return (cf. (1.5)) is

$$\tilde{R}^e_t = \frac{R^e_t}{1+\pi_t}. \tag{1.9}$$

Example 1.9 *(Real returns) With* $R_t = 0.08$ *and* $\pi_t = 0.05$*, the real return is* $1.08/1.05 - 1 \approx 0.029$*. Also, with* $R^e_t = 0.07$*, the excess real return is* $0.07/1.05 \approx 0.067$.

It is clear that the real excess return (1.9) is less affected by inflation than the real net return (1.8). (Actually, for log returns the real excess log return is unaffected by inflation.) The reason is straightforward: while inflation reduces the real value of the long position, it also reduces the real value of the short position. In practice, many investors use the traditional excess return (R^e_t, not $\tilde{R}^e_t$) as a proxy for real excess returns.

1.1.4 Descriptive Statistics of Asset Returns

The properties of returns in a sample are often summarised by the mean, standard deviation, the Sharpe ratio (mean/std of excess returns) and the coefficients from a linear regression (see below).

Remark 1.10 *(On notation) Mean returns are denoted* $\mathrm{E}\, R$ *or* μ*. (Subscripts to indicate the asset are used when needed.) An expression like* $\mathrm{E}\, x^2$ *means the expected value of* x^2 *and* $\mathrm{E}\, xy$ *is the expectation of the product* xy*. Variances are denoted* σ^2 *or* $\mathrm{Var}(R)$ *and the standard deviations* σ *or* $\mathrm{Std}(R)$*. Covariances are denoted* σ_{ij} *or* $\mathrm{Cov}(R_i, R_i)$*. Clearly,* σ_{ii} *is the same as the variance.*

The *scaling of returns* (for instance, in percentages) can often cause confusion. Let R_{it} be the net return with mean μ, standard deviation σ and covariance with asset j σ_{ij}. When you work with percentage returns, $100R_{it}$, then

$$100R_{it} \text{ has the } \begin{array}{ll} \text{mean:} & 100\mu \\ \text{variance:} & 100^2\sigma^2 \\ \text{standard deviation} & 100\sigma \\ \text{covariance with } 100R_{jt} & 100^2\sigma_{ij} \end{array} \tag{1.10}$$

Notice that the mean and standard deviation are scaled by 100, but the variance and covariance are scaled by $10,000$. This can easily 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). Also, in a regression, $\tilde{R}_{it} = \alpha + \beta \tilde{R}_{jt} + \varepsilon_{it}$, the slope is unaffected, but the intercept is scaled by 100.

It is a common convention to *annualise return statistics* when reporting the results. If the return data is for a $1/k$-year horizon (for instance, $k = 12$ for monthly data), then we typically annualise as

$$\begin{array}{ll} \text{mean:} & k\mu \\ \text{variance:} & k\sigma^2 \\ \text{standard deviation} & \sqrt{k}\sigma \\ \text{covariance with } R_j & k\sigma_{ij}. \end{array} \tag{1.11}$$

For daily data use $k = 252$ (the approximate number of trading days per year) and for weekly data $k = 52$. Also, the results from a linear regression are annualised by multiplying the intercept by k (since it is a mean), but not changing the slope coefficient (since it is a covariance divided by a variance). The convention in (1.11) is based on the idea that returns are almost iid (see below for details). It is probably advisable to annualise only at the very last stage of the computations.

Example 1.11 *(Annualisation) If the monthly average return is 0.67% and the monthly standard deviation is 2.89%, then the annualised values are 8% and 10%, respectively.*

The expected excess return, $\mathrm{E}\, R_i^e$ or μ_i^e, is often called a *risk premium* since it measures the expected return of taking risk (of holding asset i) minus the return of a risk-free asset. The *Sharpe ratio* is

$$SR = \mu^e/\sigma, \tag{1.12}$$

where (μ^e, σ) indicate the mean and standard deviations of the excess returns. The SR can be interpreted as a reward/risk ratio. Typically, a high Sharpe ratio is considered favourable.

Example 1.12 *(Risk premium and Sharpe ratio) If* $(\mu^e, \sigma) = (0.1, 0.5)$*, then Sharpe ratio is 0.2.*

The so-called "market model" is a regression of an asset's excess return on the excess return on the market index (R_{mt}^e)

$$R_t^e = \alpha + \beta R_{mt}^e + u_t. \tag{1.13}$$

A slope coefficient $\beta > 1$ indicates that the asset is strongly pro-cyclical (moves more than proportionally with the market), whereas $0 < \beta < 1$ indicates a weaker pro-cyclicality. This is sometimes referred to as "cyclical" and "defensive" assets, respectively. $\beta < 0$ indicates counter-cyclicality, but such assets are rare. The α is often interpreted as an abnormal excess return (see the chapters on CAPM and performance measures). See Table 1.1 for an example.

	Small growth	Small value	Large growth	Large value	Equity market
mean (ann.)	7.31	11.13	10.30	9.69	9.17
std (ann.)	23.39	20.00	15.99	18.50	15.62
SR (ann.)	0.31	0.56	0.64	0.52	0.59
α (ann.)	−4.46	1.44	1.16	0.43	0.00
β	1.28	1.06	1.00	1.01	1.00

Table 1.1: Means and std of asset class returns, US, monthly excess returns (%), 1985:01-2024:12. The mean and α are annualised by 12, the standard deviation by $\sqrt{12}$, and the Sharpe ratio is the ratio of the annualised mean and standard deviation.

Remark 1.13 *(Motivation of the convention in (1.11)*) Suppose we have semi-annual data. Notice that an annual return would be $P_t/P_{t-2} - 1 \approx R_t + R_{t-1}$. If returns are iid (in particular, the same mean and variance across time and also uncorrelated across time), then to a first approximation, the expected annual return is* $\mathrm{E}(R_t + R_{t-1}) = 2\,\mathrm{E}\,R_t$. *Similarly, the variance is* $\mathrm{Var}(R_t + R_{t-1}) = 2\,\mathrm{Var}(R_t)$, *which implies* $\mathrm{Std}(R_t + R_{t-1}) = \sqrt{2}\,\mathrm{Std}(R_t)$. *Similarly, if* $\mathrm{Cov}(R_{it}, R_{j,t-1}) = 0$, *then* $\mathrm{Cov}(R_{it} + R_{i,t-1}, R_{jt} + R_{j,t-1}) = 2\sigma_{ij}$.

1.1.5 Cumulating Returns

If an investment in period $t = 0$ equals V_0, then its value in t is

$$V_t = V_0(1 + R_1)(1 + R_2)\dots(1 + R_t), \tag{1.14}$$

where all subscripts refer to time periods and R_τ is the return on your portfolio in period τ. This expression assumes that all dividends have been reinvested in the same asset, making V_t a *total return index*. We can clearly write this on recursive form as

$$V_t = V_{t-1}(1 + R_t). \tag{1.15}$$

Example 1.14 *(Reinvesting dividends) Suppose we initially owned 500 shares and that the price changed from 100 to 108 and we got 2 in dividends (per share), as in*

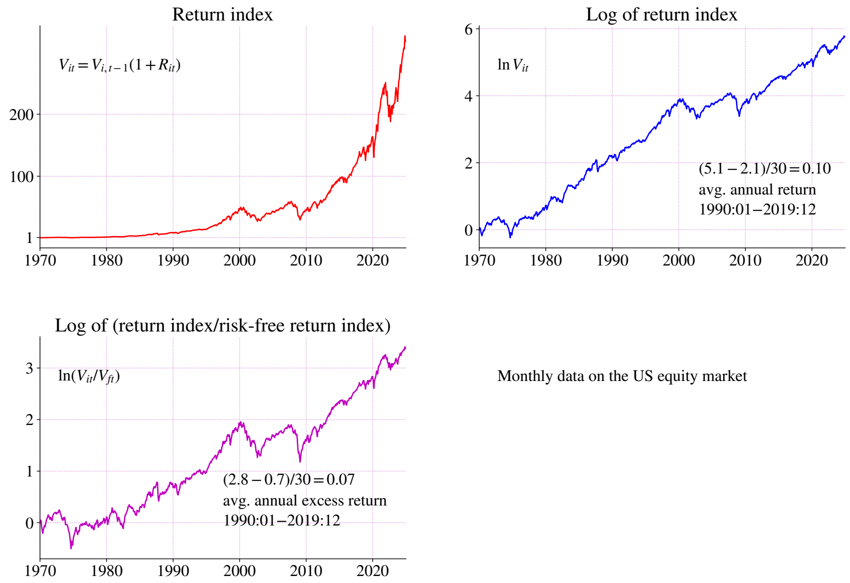


Figure 1.2: Cumulating returns

Example 1.4. Then, we would buy $500 \times 2/108 \approx 9.26$ *extra shares (assuming they are divisable, otherwise 9). In any case, our new position is worth* $509.26 \times 108 \approx 55,000$*, which is 10% more than the initial value.*

Empirical Example 1.15 *Figure 1.2 shows the cumulated return of a U.S. equity market index.*

Example 1.16 *With net returns for three time periods* $(R_1, R_2, R_3) = (0.2, -0.35, 0.25)$*, we get portfolio values* $(1.2, 0.78, 0.975)$ *for period* $1-3$ *(assuming* $V_0 = 1$*).*

Remark 1.17 *(Adjusted closing price) The adjusted closing price of an asset is an index calculated as (1.15) where* R_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* $R_t = V_t/V_{t-1} - 1$*, without having to handle dividend payments separately.*

Unfortunately, excess returns cannot be cumulated directly. Instead, you need to cumulate the net return R_t and and the risk-free return R_{ft} separately (as in (1.15)) and then form the difference

$$V_t^e = V_t - V_{ft}. \tag{1.16}$$

Sometimes the ratio V_t / V_{ft} is a preferred way of illustrating the performance of the two assets.

1.1.6 Cumulating Logarithmic Returns*

Similarly, the log value can be calculated as

$$\begin{aligned} \ln V_t &= \ln V_0 + r_1 + r_2 \ldots + r_t, \text{ so} && (1.17)\\ &= \ln V_{t-1} + r_t. && (1.18) \end{aligned}$$

You *can* cumulate excess log returns (because it is just summing). Since the initial positions are equal ($V_0 = V_{f,0}$) we have

$$\begin{aligned} \ln(V_t / V_{ft}) &= (r_1 + r_2 + \ldots + r_t) - \left(r_{f1} + r_{f2} \ldots + r_{ft}\right) && (1.19)\\ &= r_1^e + r_2^e + \ldots + r_t^e, \text{ so} && (1.20)\\ &= \ln(V_{t-1} / V_{f,t-1}) + r_t^e, && (1.21) \end{aligned}$$

starting from $\ln(V_0 / V_{f0}) = 0$. Notice that the exponential function of this gives the ratio V_t / V_{ft} (not the difference). Again, see Figure 1.2 for an illustration

1.2 Portfolio Returns

Remark 1.18 *(On notation) These notes use $\Sigma_{i=1}^n x_i$ to denote the sum $x_1 + \ldots + x_n$. (In the running text, it might happen that this is sometimes written as just $\Sigma_i x_i$.) Note: Σ may also denote a variance-covariance matrix. The distinction should be clear from the context.*

1.2.1 Portfolio Return: Definition

Let R_i represent the return on asset i over a given time period (the time subscript is omitted for convenience). The return on a portfolio (R_p) with the portfolio weights $w_1, w_2, ..., w_n$ is

$$R_p = \textstyle\sum_{i=1}^n w_i R_i, \text{ with } \sum_{i=1}^n w_i = 1. \tag{1.22}$$

Using vectors, this can also be written

$$R_p = w'R, \tag{1.23}$$

where w is an n-vector of weights and R an n-vector of asset returns.

Clearly, one of the assets in (1.22)–(1.23) could be risk-free with return R_f. However, in this case we will typically choose to consider n risky assets and the risk-free (in total, $n + 1$) and write the portfolio return as

$$R_p = v'R + (1 - \mathbf{1}'v)R_f \tag{1.24}$$

$$= v'R^e + R_f, \tag{1.25}$$

where v are the weights on the risky assets and $1-\mathbf{1}'v$, that is, $1-\Sigma_{i=1}^n v_i$, the weight on the risk-free asset. This automatically imposes the condition that the weights on *all* assets sum to one.

Example 1.19 *(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 1.20 *(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* $P_{2,t-1} = 8, P_{2,t} = 8.4$. *Assume 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* $R_p = (218 - 200)/200 = 0.09$. *This is the same as in Example 1.19 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.*

1.2.2 Portfolio Return with Short Positions

The portfolio weights in (1.22) 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% in a particular asset. In contrast, hedge funds have very few limits.

Remark 1.21 *(Short selling) How can we short sell an asset? Borrow the asset (for a fee and typically against collateral) and sell it. If there are derivatives on the asset, then we do not need to borrow it: just issue a futures/option.*

Example 1.22 *(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%*.*

1.2.3 Zero-Cost Portfolios*

A *zero-cost "portfolio"* (also called an arbitrage portfolio) means that the investor shortens some assets (perhaps borrows) in order to invest in 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 in a "long" portfolio and denote the weights by w_i^L, and and a "short" portfolio with weights w_i^S. Clearly, $w_i^L \geq 0$ and $w_i^S \geq 0$ and when one of them is positive then the other is zero.

Example 1.23 *(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

$$R_p^L = \textstyle\sum_{i=1}^n w_i^L R_i \tag{1.26}$$

$$R_p^S = \textstyle\sum_{i=1}^n w_i^S R_i, \tag{1.27}$$

where all subscripts refer to different assets (and the subscripts for time are suppressed).

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

$$R_p^e = R_p^L - R_p^S = \textstyle\sum_{i=1}^n (w_i^L - w_i^S) R_i. \tag{1.28}$$

Conversely, the traditional excess return of an asset (1.5) is the return of a zero cost portfolio: a long position in the asset and a short position in the risk-free asset.

Example 1.24 *(Excess return of a zero-cost portfolio) If the returns of the assets in Example 1.23 are* $R_1 = 10\%$, $R_2 = -1\%$ *and* $R_3 = 2\%$*, then the excess return is*

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

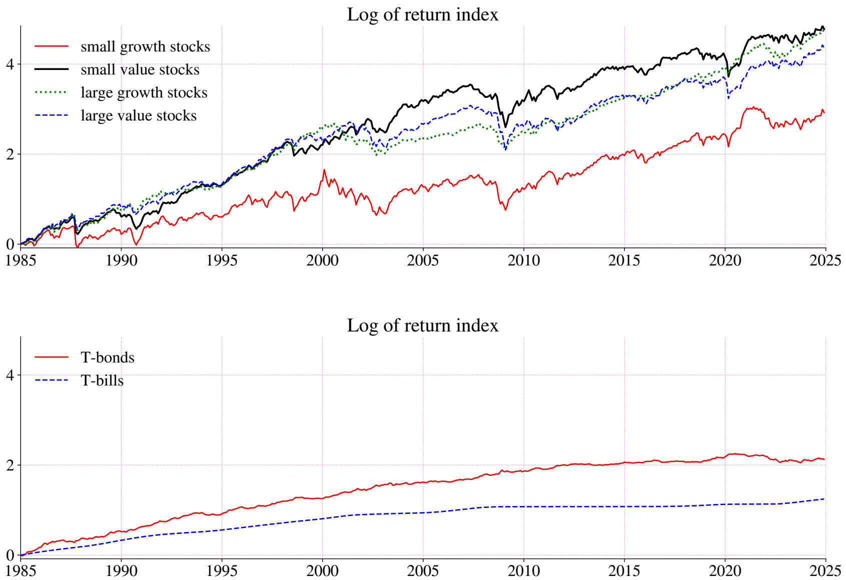


Figure 1.3: Performance of US equity and fixed income

Remark 1.25 *(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).*

1.2.4 Trading Costs

As a an investor you typically pay a *commission* to the broker. In addition, the price depends on whether you are buying (high price, the ask price) or selling (low price, the bid price).

Notice that portfolio weights shift as a result of returns (price changes). After returns have been realized, the new portfolio weight on asset i is

$$w_{it}(1 + R_{it})/(1 + R_{pt}), \tag{1.29}$$

where w_{it} was the initial weight. If $w_{i,t+1}$ is the desired weight going forward, the trading need for the portfolio is the absolute value of the difference to (1.29)

$$\sum_{i=1}^{n} |w_{i,t+1} - w_{it}(1 + R_{it})/(1 + R_{pt})|. \tag{1.30}$$

Example 1.26 *(Trading need) For Example 1.19, (1.29) gives* $0.8 \times 1.1/1.09 \approx 0.807$ *and* $0.2 \times 1.05/1.09 \approx 0.193$*, respectively for the two assets. If the aim is to keep the weights fixed, then we need to sell off (buy) approximately* 0.007 *of asset 1 (2), so the total trading need is approximately* 0.014.

Remark 1.27 *(Trading costs*) Suppose* 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.*

1.3 Asset Classes

Many investors and asset managers choose to focus on asset classes, rather than on individual assets. This approach helps average out idiosyncratic (for instance, firm specific) noise and focuses attention to the macroeconomic perspective.

Empirical Example 1.28 *Table 1.2 illustrates the return distributions for different U.S. asset classes. There are distinct differences between small and large firms and between growth and value firms. However, the most pronounced difference is between equity and bonds (the latter have much less volatility and often lower returns). 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.*

	Small growth	Small value	Large growth	Large value	Bonds	T-bills
mean	0.87	1.19	1.12	1.07	0.45	0.26
std	6.74	5.76	4.62	5.34	1.40	0.21
min	−32.48	−28.09	−23.22	−27.23	−4.39	0.00
max	28.09	19.54	14.47	18.17	5.31	0.79
market corr	0.86	0.82	0.97	0.85	0.00	0.03
beta	1.28	1.05	1.00	1.01	0.00	0.00

Table 1.2: Descriptive statistics of asset classes, US, monthly returns (%), 1985:01-2024:12. The beta is the slope coefficient from regressing the asset on the market return.

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 in the secondary market can be done on an *exchange* (NYSE, Tokyo, EuroNext, Nasdaq, London, Shanghai, HK, CBOE, CME, etc), an *electronic platform* (EBS, Reuters), or *over the counter* (OTC).

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

Alternative asset classes (for instance, hedge funds, infrastructure, private equity) have gained interest over the last decade, especially among institutional long-run investors (wealthy individuals, endowments, some pension funds).

Remark 1.29 *(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)*

	6th		5th		4th		3rd		2nd		1st	
2005	SG	0	B	3	TB	3	LG	4	SV	10	LV	14
2006	B	3	TB	5	SG	9	LG	11	LV	22	SV	22
2007	SV	-14	LV	-2	TB	5	SG	6	B	9	LG	13
2008	SG	-41	LV	-39	LG	-34	SV	-34	TB	2	B	14
2009	B	-4	TB	0	LV	18	SV	30	LG	31	SG	37
2010	TB	0	B	6	LV	7	LG	15	SV	27	SG	29
2011	LV	-11	SV	-8	SG	-6	TB	0	LG	4	B	10
2012	TB	0	B	2	SG	15	LG	15	SV	21	LV	29
2013	B	-3	TB	0	LG	33	LV	40	SV	43	SG	45
2014	TB	0	SV	4	SG	5	B	5	LV	12	LG	14
2015	SV	-10	LV	-8	SG	-3	TB	0	B	1	LG	4
2016	TB	0	B	1	SG	8	LG	9	LV	26	SV	37
2017	TB	1	B	2	SV	9	LV	18	SG	25	LG	29
2018	LV	-15	SV	-13	SG	-8	LG	0	B	1	TB	2
2019	TB	2	B	7	SV	15	LV	28	SG	30	LG	34
2020	LV	-3	TB	0	SV	4	B	8	LG	36	SG	57
2021	B	-2	TB	0	SG	3	LG	25	LV	37	SV	42
2022	SG	-28	LG	-26	B	-12	SV	-6	TB	1	LV	4
2023	B	4	TB	5	SV	13	LV	14	SG	16	LG	38
2024	B	1	TB	5	SV	9	SG	18	LV	21	LG	30

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.

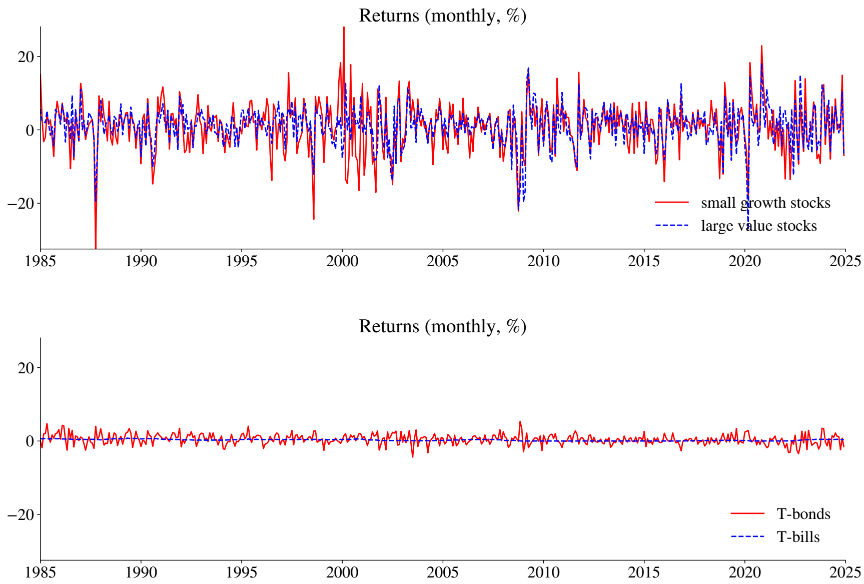


Figure 1.4: Performance of US equity and fixed income

Chapter 2

The Basics of Portfolio Choice

This is the first of several chapters on portfolio choice, introducing two basic concepts: *(1)* how to mix the risky assets with a risk-free asset to handle the overall risk level (leverage); and *(2)* how to mix various risky assets to average out volatility (diversification). This chapter introduces each of these topics. Later chapters will put them together in a unified framework and also discuss optimal portfolios.

2.1 Expected Portfolio Return and Variance

This technical section summarizes how beliefs about expected returns (μ) of the investable assets and their variance-covariance matrix (Σ) can be combined with portfolio weights to calculate the implied beliefs about the portfolio returns. These beliefs should be interpreted as representing those of the investor, conditional on the information available at the time of the investment. In later chapters, we will extend this approach to find optimal weights.

Remark 2.1 *(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*

$$\begin{aligned}
\mathrm{E}(w_1 R_1 + w_2 R_2) &= w_1 \mu_1 + w_2 \mu_2, \text{ and} \\
\mathrm{Var}(w_1 R_1 + w_2 R_2) &= w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2 w_1 w_2 \sigma_{12},
\end{aligned}$$

where $\mu_i = \mathrm{E}\, R_i$, $\sigma_{ij} = \mathrm{Cov}(R_i, R_j)$, and $\sigma_i^2 = \mathrm{Var}(R_i)$.

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

$$\begin{aligned} \mathrm{E}\, R_p &= \textstyle\sum_{i=1}^{n} w_i \mu_i \quad (2.1) \\ &= w'\mu, \quad (2.2) \end{aligned}$$

where w is the n-vector of portfolio weights and μ is a corresponding vector of expected asset returns.

The variance of a portfolio return is

$$\begin{aligned} \mathrm{Var}(R_p) &= \textstyle\sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} w_i w_j \sigma_{ij} \quad (2.3) \\ &= w' \Sigma w, \quad (2.4) \end{aligned}$$

where Σ is the $n \times n$ variance-covariance matrix of the returns.

Remark 2.2 *($n = 2$) With two assets,* $\mathrm{E}\, R_p = w_1\mu_1 + w_2\mu_2$, $\mathrm{Var}(R_p) = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_{12}$ *and* $\mathrm{Cov}(R_q, R_p) = v_1w_1\sigma_1^2 + v_2w_2\sigma_2^2 + (v_1w_2 + v_2w_1)\sigma_{12}$.

Example 2.3 *(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:* $\mu = \begin{bmatrix} 9 \\ 6 \end{bmatrix} /100$ *and* $\Sigma = \begin{bmatrix} 256 & 96 \\ 96 & 144 \end{bmatrix} /100^2$. *This gives*

$$\begin{aligned} \mathrm{E}\, R_p &= \begin{bmatrix} 0.8 & 0.2 \end{bmatrix} \begin{bmatrix} 9 \\ 6 \end{bmatrix} \frac{1}{100} = 0.084, \\ \mathrm{Var}(R_p) &= \begin{bmatrix} 0.8 & 0.2 \end{bmatrix} \begin{bmatrix} 256 & 96 \\ 96 & 144 \end{bmatrix} \frac{1}{100^2} \begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix} \approx 0.020, \text{ and} \\ \mathrm{Std}(R_p) &\approx 0.142. \end{aligned}$$

More details and examples are found in the statistics appendix.

2.2 Leverage

2.2.1 A Portfolio of a Single Risky Asset and a Risk-free Asset

Suppose you can only invest in a risky asset (with return R) and a risk-free (with return R_f). The risky asset could represent the (equity) market portfolio. To observe the effect of the portfolio choice on the mean and the volatility, notice that

$$R_p = vR + (1-v)R_f, \text{ so} \tag{2.5}$$

$$\mathrm{E}\,R_p = v\mu + (1-v)R_f \text{ and} \tag{2.6}$$

$$\mathrm{Std}(R_p) = |v|\sigma, \tag{2.7}$$

where we use (μ and σ) as short hand notation for the mean and standard deviation of the risky asset.

The expected value follows from $\mathrm{E}\,R_f = R_f$ as the risk-free rate is known. Similarly, the standard deviation follows from $\mathrm{Var}(R_p) = v^2\sigma^2$, since $\mathrm{Var}(R_f) = 0$ (the risk-free 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 risk-free rate, then we should typically use a maturity that corresponds to the investment horizon. Often a floating overnight rate is used instead, but that is (strictly speaking) not risk-free for investment horizons of more than one day. Still, the uncertainly might be so small that it can be used as an approximation.

How much to put in the risky asset is a matter of *leverage*, and v is often called the *leverage ratio*. This equals the investment in risky assets divided by our total capital.

Example 2.4 *(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 risk-free)*	0.5	0	-1	2
Sum	1	1	1	1

Portfolio A*: your capital is* 200*, invest* 100 *in risky assets and* 100 *in risk-free;* Portfolio B*: your capital is* 200*, invest* 200 *in risky assets and* 0 *in risk-free;* Portfolio C*: your capital is* 200*, invest* 400 *in risky assets and* –200 *in risk-free (borrow 200 = short position in risk-free). Portfolio D: short-sell the risky asset for* 100 *and put* 200 *in the risk-free.*

Remark 2.5 *(Assuming that R and R_f do not depend on v) These notes 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.*

The mean and the standard deviation in (2.6)–(2.7) are both scaled by the leverage ratio (v). Notice that taking on leverage (borrowing to invest in the risky asset) typically is a way to increase the expected return of the portfolio, but at the cost of increasing the risk.

Empirical Example 2.6 *Figure 2.1 shows the effect on the log cumulated excess returns from holding a leveraged position in equity. The LIBOR rate (London Interbank Offered Rate) is, of course, not entirely without variation in this figure, thus, the result in (2.7) applies only approximately. However, for each separate 1-month investment horizon, the LIBOR rate is known in advance and thus risk-free. More recently, risk-free rates are often proxied by floating overnight rates.*

Example 2.7 *With* $(\mu, \sigma) = (9.5\%, 8\%)$ *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 (2.6)–(2.7) to get a relation between portfolio mean and standard deviation as

$$\mathrm{E}\, R_p = R_f + SR \times \mathrm{Std}(R_p), \tag{2.8}$$

where the slope is $SR = \mu^e/\sigma$ (the Sharpe ratio of the risky asset). This shows that the average portfolio return is linearly related to its standard deviation. (For $v < 0$, the relation is also linear, but with the slope $-SR$.) Figure 2.2 illustrates.

2.3 Diversification

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

Recall that the variance of a portfolio return is

$$\mathrm{Var}(R_p) = w'\Sigma w, \tag{2.9}$$

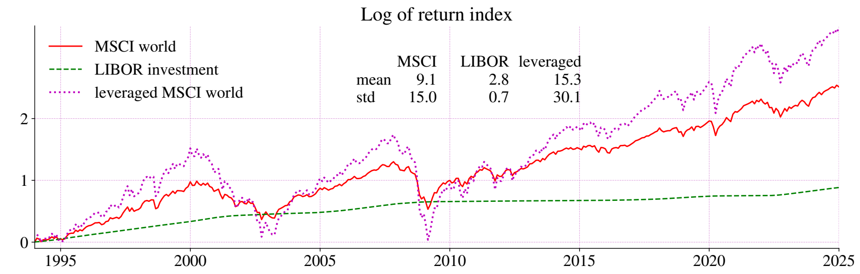


Figure 2.1: The effect of leverage on the portfolio performance

where w is the vector of portfolio weights and Σ the variance-covariance matrix. For instance, with two assets we have

$$\mathrm{Var}(R_p) = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_{12}, \tag{2.10}$$

where w_i is the portfolio weight on asset i, σ_i^2 is the variance of asset i and σ_{ij} is the covariance of assets i and j.

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 (2.10)). Denote the correlation by ρ and write as (since $\sigma_{12} = \rho\sigma_1\sigma_2$)

$$\begin{aligned}\mathrm{Var}(R_p) &= \sigma_1^2/4 + \sigma_2^2/4 + \rho\sigma_1\sigma_2/2\\ &= \sigma^2(1+\rho)/2 \text{ if } \sigma_1 = \sigma_2 = \sigma,\end{aligned} \tag{2.11}$$

where the second equality assumes that both assets have the same standard deviation.

If the assets are uncorrelated ($\rho = 0$), then the variance of this portfolio is half that of the assets—which demonstrates the importance of diversification. This effect is even stronger when the correlation is negative: with $\rho = -1$ the portfolio variance is actually zero, which we call *hedging*. In contrast, with a high correlation, the benefit from diversification is smaller (and zero when the correlation is perfect, $\rho = 1$). See Figure 2.3 for an illustration.

Example 2.8 *(Diversification) If* $\sigma = 16\%$ *(so* $\sigma^2 = 256/100^2$*) and* $\rho = 0.5$*, then (2.11) gives* $\mathrm{Var}(R_p) = 192/100^2$ *and thus* $\mathrm{Std}(R_p) \approx 13.9\%$.

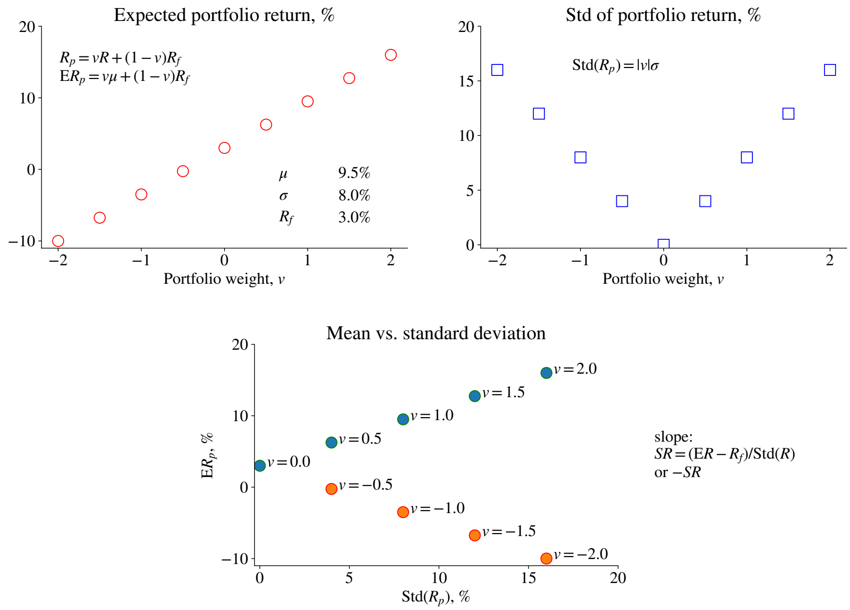


Figure 2.2: The effect of leverage on the mean and volatility of the portfolio return

Empirical Example 2.9 *Table 2.1 provides an empirical example of the correlations between major asset classes.*

2.3.2 Diversification: The Number of Assets

In order to see the importance of mixing many assets in the portfolio, we will consider equally weighted portfolios of n assets ($w_i = 1/n$), to focus on the basic idea. Clearly, there are other (not equally weighted) portfolios with even lower variance.

The variance of an equally weighted ($w_i = 1/n$ so $w_i^2 = 1/n^2$) portfolio is

$$\mathrm{Var}(R_p) = (\overline{\sigma}^2 - \overline{\sigma}_{ij})/n + \overline{\sigma}_{ij}, \tag{2.12}$$

where $\overline{\sigma}^2$ is the average variance (average across the n assets) and $\overline{\sigma}_{ij}$ is the average covariance of two returns. Both can be treated as constants if we pick assets randomly. In case the assets are uncorrelated, (2.12) shows that the portfolio variance goes to zero as the number of assets (included in the portfolio) goes to infinity. More realistically, $\overline{\sigma}_{ij}$ is positive. When the portfolio includes many assets, then the

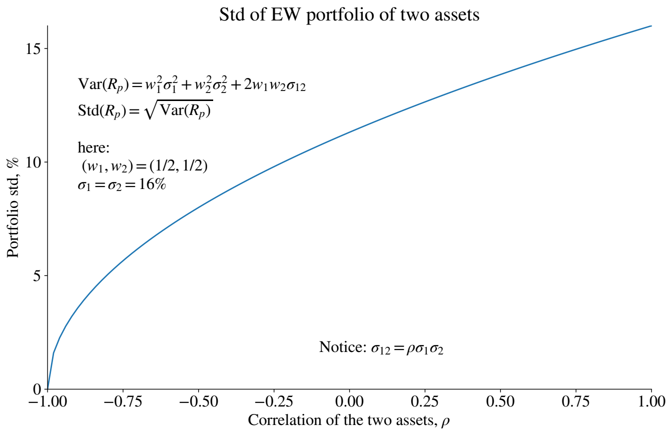


Figure 2.3: Effect of correlation on the diversification benefits

	Small growth	Small value	Large growth	Large value	Bonds	T-bills
Small growth	1.00	0.86	0.80	0.69	−0.09	−0.03
Small value	0.86	1.00	0.71	0.85	−0.10	−0.03
Large growth	0.80	0.71	1.00	0.76	0.05	0.04
Large value	0.69	0.85	0.76	1.00	−0.06	0.04
Bonds	−0.09	−0.10	0.05	−0.06	1.00	0.20
T-bills	−0.03	−0.03	0.04	0.04	0.20	1.00

Table 2.1: Correlations of asset class returns, US, monthly returns, 1985:01-2024:12

average covariance dominates. In the limit (as n goes to infinity), only this non-diversifiable risk matters, $\overline{\sigma}_{ij}$. See Elton, Gruber, Brown, and Goetzmann (2014) 4 for a more detailed discussion.

Example 2.10 *(Variance of portfolio return) With* $\overline{\sigma}^2 = 256/100^2$ *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* $(16\%, 13.9\%, 13.1\%)$.

Empirical Example 2.11 *Figure 2.4 shows an empirical example of what diversification implies. 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 2.5 suggests that the diversification benefits are not constant across time.*

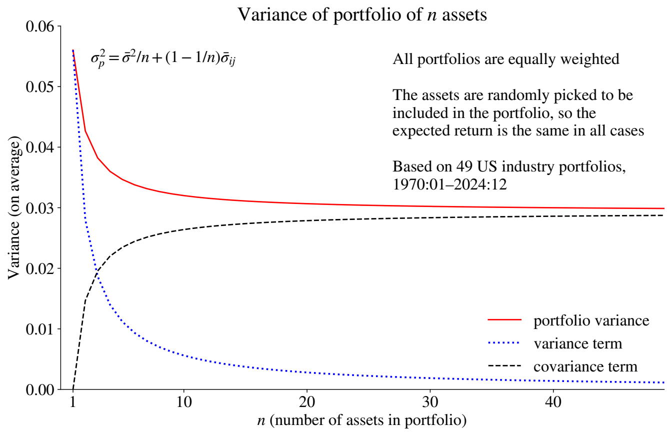


Figure 2.4: Effect of diversification

Proof of (2.12). Note that $\text{Var}(R_p) = (\mathbf{1}/n)'\Sigma(\mathbf{1}/n)$, where $\mathbf{1}$ is a vector of ones. This is just summing the elements in Σ and dividing by n^2. In this sum, there are n variances and $n(n-1)$ covariances. We can thus write the variance as

$$\begin{aligned}\text{Var}(R_p) &= \frac{1}{n}\sum\nolimits_{i=1}^{n}\frac{\sigma_i^2}{n} + \frac{n-1}{n}\sum\nolimits_{i=1}^{n}\sum\nolimits_{j=1,j\neq i}^{n}\frac{\sigma_{ij}}{n(n-1)} \\ &= \overline{\sigma}^2/n + \overline{\sigma}_{ij}(n-1)/n,\end{aligned}$$

which can be rearranged as (2.12). □

Remark 2.12 *(On negative covariances in (2.12)*) 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 that are, on average, negatively correlated with each other. In (2.12) this manifests itself in that $\overline{\sigma}_{ij} < 0$ would give a negative portfolio variance as n increases.*

2.4 Covariances Do Matter

This section will (once again) illustrate the importance of covariances for the portfolio variance, but from another perspective. We relax the assumption of an equal weights, but consider only small changes to an existing portfolio.

Suppose we are initially invested in a portfolio p (with portfolio weights of the risky assets in the vector v risky assets and $1 - v'\mathbf{1}$ in the risk-free). We now consider

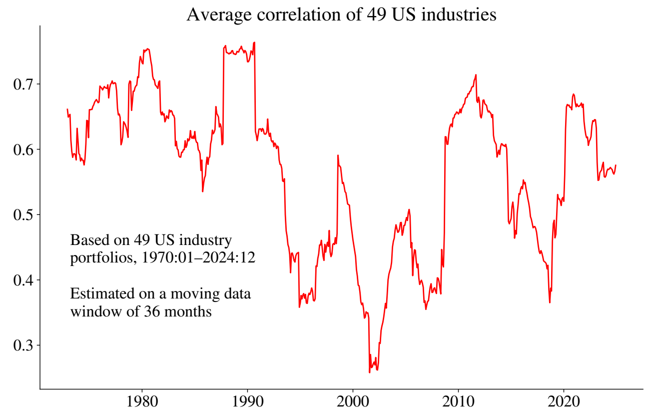


Figure 2.5: Time-varying correlations

a small increase (δ) of the portfolio weight of asset i financed by borrowing at the risk-free rate. The portfolio return of the new portfolio (q) would then be

$$R_q = R_p + \delta R_i^e, \tag{2.13}$$

The *incremental return*, δR_i^e, is just δ times the excess return on the asset i. This is straightforward since we have increased the exposure to asset i and financed it with borrowing at the risk-free rate.

The portfolio variance is

$$\sigma_q^2 = \sigma_p^2 + \underbrace{\delta^2\sigma_i^2 + 2\delta\sigma_{ip}}_{\text{incremental variance}}, \tag{2.14}$$

where σ_{ip} is the covariance of our portfolio p with asset i.

The *incremental variance* is $\delta^2\sigma_i^2 + 2\delta\sigma_{ip}$, so it depends on the variance of asset i and on how it correlates with our current portfolio p. For small values of δ (say, $\delta = 5\%$) the *covariance effect might dominate* (since δ^2 decreases very quickly). Conversely, adding a small amount of an uncorrelated asset ($\sigma_{ip} = 0$) to your portfolio does not change the portfolio variance much at all. See Figure 2.6 for an illustration.

Example 2.13 *(of (2.14)) The easiest case is when σ_p and σ_i both equal 1, so σ_{ip} equals the correlation ρ. Then, the incremental variance is $\delta^2 + 2\delta\rho$. For $\delta = 0.05$ we have* $0.0025 + 0.08$ *when* $\rho = 0.8$ *so the covariance effect is 32 times larger than*

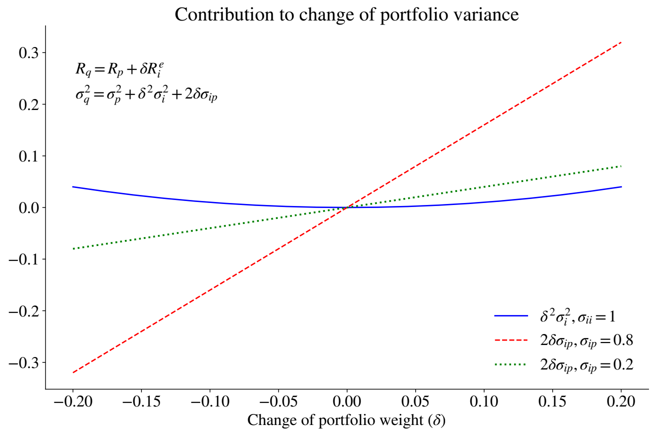


Figure 2.6: The effect of a portfolio change on the variance

the effect of $\delta^2\sigma_i^2$. *Conversely, with* $\rho = 0$ *the covariance effect is zero. See also Figure 2.6.*

Chapter 3

The Mean-Variance Frontier

This chapter introduces the mean-variance frontier, which is an approach for summarizing what portfolio formation can achieve in terms of average returns and volatility. It will identify efficient portfolios (those on the frontier) that dominate other portfolios.

3.1 The Mean-Variance Frontier of Risky Assets

The mean-variance frontier (MVF, see Markowitz (1952)) is based on the idea that the investor seeks high average portfolio returns but dislikes portfolio return variance.

To find the mean-variance frontier, we first have to specify the n-vector of average returns of the investable assets (μ) and their variance-covariance matrix (Σ). As in earlier chapters, the means and the variance-covariance matrix should be interpreted as representing the investor's beliefs, conditional on the information available at the time of the investment.

	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 three assets in some 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, $\times 100^2$).

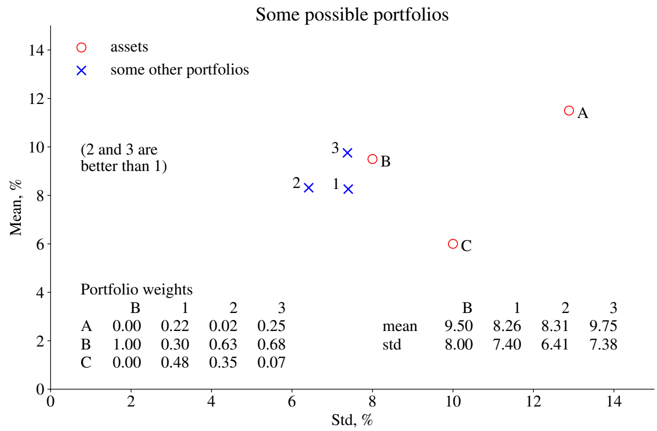


Figure 3.1: Mean vs standard deviation. The properties of the investable assets (A, B, and C) are shown in Table 3.1.

Example 3.1 *(Mean and Std of a portfolio) Table 3.1 illustrates a case with three investable assets (A, B and C). The mean returns are given in percentages; thus, 6% should be read as 0.06. In contrast, the variance-covariance matrix is given in terms of basis points (bp, where 1bp $= 1/10000$); thus, 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 illustrates the location of the investable assets (A, B and C) from Table 3.1, as well as some portfolios $(1, 2, 3)$ of them. The figure has the *standard deviation* on the horizontal axis and the *expected return* on the vertical axis. It is reasonable to think that portfolio 3 is better than B (lower volatility and higher expected returns) and also that portfolios 2 and 3 are better than portfolio 1 (lower volatiltity and higher expected returns, respectively). The mean-variance frontier extends this logic by considering all possible portfolios based on the same investable assets.

To calculate a point on the MVF, 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

$$\min_{w_i} \text{Var}(R_p) \text{ subject to} \tag{3.1}$$

$$\text{E}\, R_p = \mu^* \text{ and } \textstyle\sum_{i=1}^{n} w_i = 1.$$

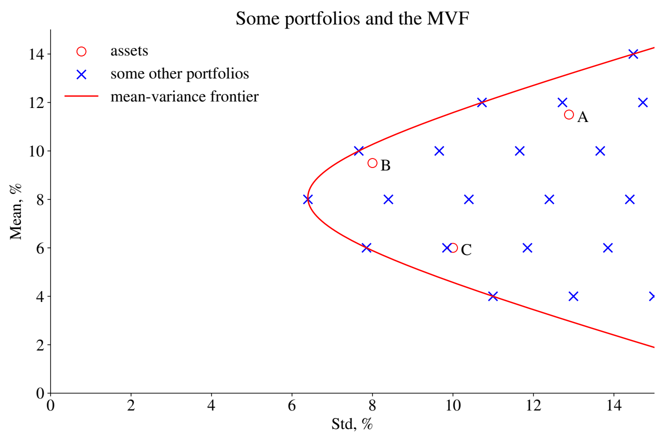


Figure 3.2: Some portfolios and the mean-variance frontier. The properties of the investable assets (A, B, and C) are shown in Table 3.1.

The solution can be found with numerical methods or linear algebra, as shown below.

Remark 3.2 *(Portfolio average and variance) 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. The portfolio mean and variance can then be calculated as* $\mathrm{E}\, R_p = w'\mu$ *and* $\mathrm{Var}(R_p) = w'\Sigma w$.

The whole mean-variance frontier is generated by solving this problem for different values of the expected return, μ^*. See Figure 3.2 for an example and comparison with some other portfolios with the same expected return. The *efficient frontier* (EF) is the upper leg of the curve. Reasonably, a portfolio on the lower leg is dominated by one on the upper leg at the same volatility (since it has a higher expected return). Notice that there are no portfolios (based on the given investable assets and their assumed properties μ and Σ) above or to the left of the efficient frontier.

Remark 3.3 *(How many different portfolios are there with $\mathrm{E}\, R_p = \mu^*$?) With two assets, we require $w\mu_1 + (1-w)\mu_2 = \mu^*$ and there is only one choice of w that satisfies 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.*

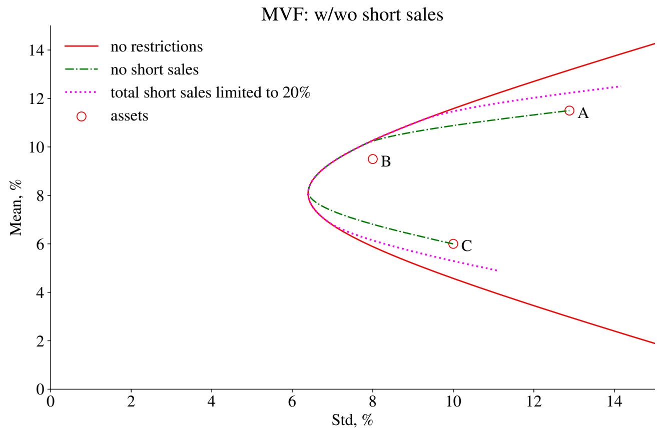


Figure 3.3: Mean-variance frontiers with restrictions. The properties of the investable assets (A, B, and C) are shown in Table 3.1.

3.1.1 The Mean Variance Frontier with Portfolio Restrictions

There are sometimes *additional restrictions*, for instance, of no short sales

$$\text{no short sales: } w_i \geq 0, \tag{3.2}$$

In other cases, there are both lower and upper bounds on the weights

$$L_i \leq w_i \leq U_i. \tag{3.3}$$

For instance, mutual funds often have to obey $L_i = 0$ and $U_i = 0.1$.

Funds may also impose restrictions on themselves; for instance, they may allow limited short sales

$$\text{limited total short sales: } \textstyle\sum_{i=1}^{n} \min(w_i, 0) \geq Q. \tag{3.4}$$

With such restrictions we typically have to apply some explicit numerical minimization algorithm to find portfolio weights. Algorithms that solve quadratic problems are best suited. See Figures 3.3 –3.4 for an example.

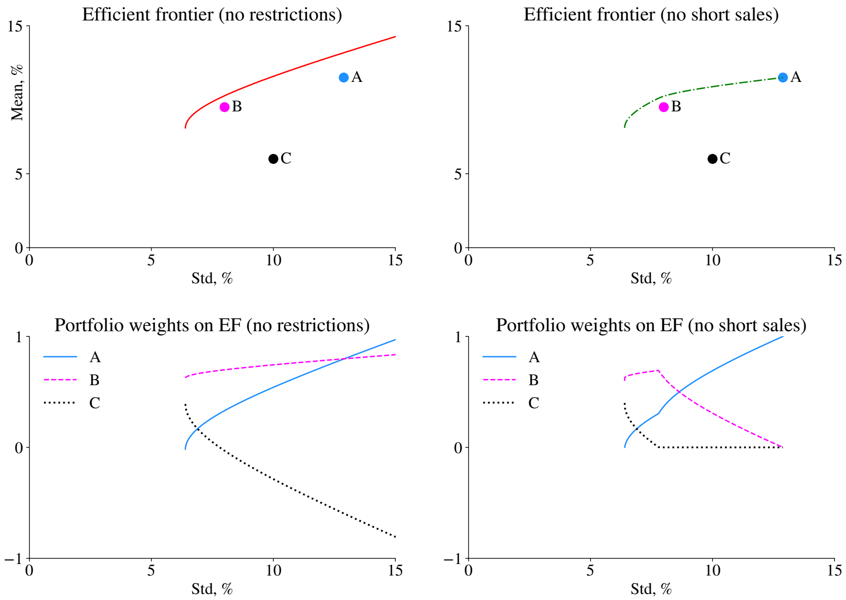


Figure 3.4: Portfolio weights on the efficient frontier. The properties of the investable assets (A, B, and C) are shown in Table 3.1.

3.1.2 The Mean Variance Frontier with Two Risky Assets

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

$$\mathrm{E}\, R_p = w\mu_1 + (1-w)\mu_2 \tag{3.5}$$

$$\mathrm{Var}(R_p) = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2w(1-w)\sigma_{12}. \tag{3.6}$$

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, all portfolios of them are on the mean-variance frontier (cf. Remark 3.3). For that reason no explicit minimization is needed. See Figure 3.5 for an example.

3.1.3 The Shape of the Mean-Variance Frontier of Risky Assets

Consider what happens when we *add assets to the investment opportunity set*. The old mean-variance frontier is, of course, still obtainable: we can always put zero weights on the new assets. In most cases, we can do better than that: the mean-variance

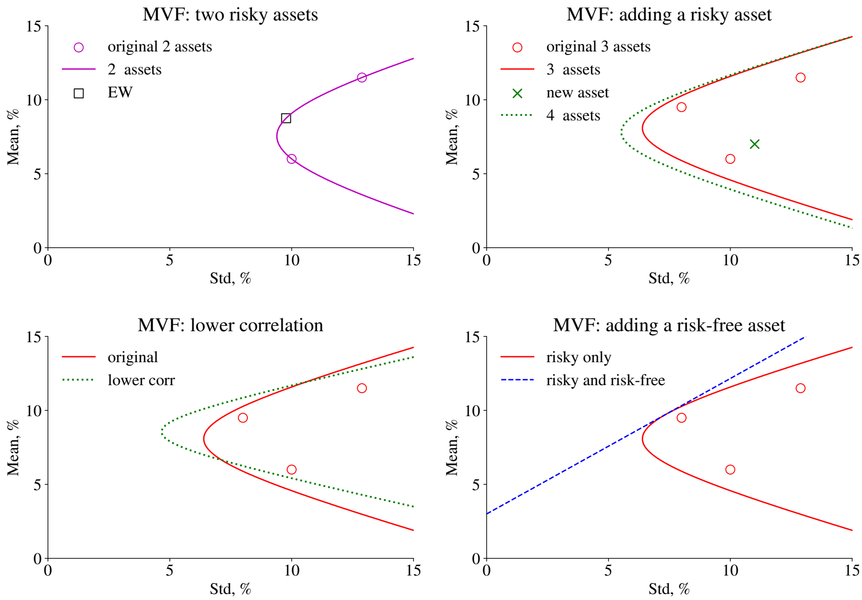


Figure 3.5: The shape of the mean-variance frontiers. The properties of the investable assets (A, B, and C) are shown in Table 3.1.

frontier is shifted to the left (lower volatility at the same expected return). See Figure 3.5 for an example. In this example the new asset is not very attractive (low average returns, high volatility), but it may be useful in a portfolio (diversification, or for shortening).

With intermediate correlations ($-1 < \rho < 1$), the mean-variance frontier is a hyperbola (see Figure 3.5). Notice that the mean–volatility trade-off improves as the correlation decreases: a lower correlation means that we get a lower portfolio standard deviation at the same expected return—at least for the efficient frontier (above the bend).

When we allow investment also in a risk-free asset (to be discussed in detail in a separate section), then the MVF becomes a straight line (again, see Figure 3.5).

Empirical Example 3.4 *Figure 3.6 shows the MVF implied by the sample means and variance-covariance matrix of 10 U.S. industry portfolios. It is therefore an ex post construction, which may (or not) be close the beliefs investors held during the sample period.*

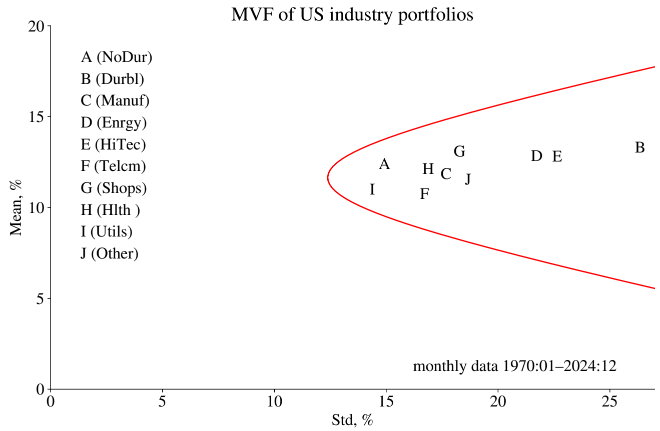


Figure 3.6: Mean-variance frontier from US industry portfolios

3.1.4 Calculating the Mean-Variance Frontier of Risky Assets

When there are no restrictions on the portfolio weights, then both numerical optimization or some simple matrix algebra can solve the optimization problem. The section demonstrates the second approach.

The minimization problem (3.1) can be written

$$\begin{aligned} &\min_w w'\Sigma w \text{ subject to} \\ &w'\mu = \mu^* \text{ and } w'\mathbf{1} = 1, \end{aligned} \tag{3.7}$$

where $\mathbf{1}$ is a vector of n ones (as many as there are assets). Again, μ and Σ summarise the beliefs of the investor, conditional on the information available at the time of the investment.

Remark 3.5 *(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 small (or large) as possible. The answer is b_{low}, b_{high}, or a value of b where $df(b)/db = 0$.*

The first order conditions are

$$\begin{bmatrix} \Sigma & \mu & \mathbf{1} \\ \mu' & 0 & 0 \\ \mathbf{1}' & 0 & 0 \end{bmatrix} \begin{bmatrix} w \\ \lambda \\ \delta \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mu^* \\ 1 \end{bmatrix}, \tag{3.8}$$

where $\mathbf{0}$ is a vector of n zeros. Solve for the vector (w, λ, δ) and extract the w vector.

Using the solution in $\sqrt{w'\Sigma w}$ gives the standard deviation of a portfolio with expected return μ^* (which should equal $w'\mu$). We can trace out the entire mean-variance frontier, by repeating this calculations for different values of the required return μ^* and then connecting the dots. In the std×mean space, the efficient frontier (the upper part) is *concave*. See, for instance, Figure 3.2.

Proof of (3.8). We set up this as a Lagrangian problem

$L = (w_1^2\sigma_{11} + w_2^2\sigma_{22} + 2w_1w_2\sigma_{12})/2 + \lambda(w_1\mu_1 + w_2\mu_2 - \mu^*) + \delta(w_1 + w_2 - 1).$ Dividing the variance by 2 does not affect the solution. (Variances are denoted σ_{ii} in order to facilitate comparison with the matrix expressions.) The first order conditions with respect to the portfolio weights are

$$\text{for } w_1 : w_1\sigma_{11} + w_2\sigma_{12} + \lambda\mu_1 + \delta = 0,$$

$$\text{for } w_2 : w_1\sigma_{12} + w_2\sigma_{22} + \lambda\mu_2 + \delta = 0.$$

Similarly, the first order conditions with respect to the Lagrange multipliers are

$$\text{for } \lambda : w_1\mu_1 + w_2\mu_2 = \mu^*,$$

$$\text{for } \delta : w_1 + w_2 = 1.$$

In matrix notation these first order conditions are

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \mu_1 & 1 \\ \sigma_{12} & \sigma_{22} & \mu_2 & 1 \\ \mu_1 & \mu_2 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \lambda \\ \delta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \mu^* \\ 1 \end{bmatrix},$$

which is (3.8). □

3.1.5 Further Properties of the MVF of Risky Assets*

Remark 3.6 states that the weights for a portfolio on the MVF of risky assets (at a given required return μ^*) can be solved as

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

where $(\tilde{\lambda}, \tilde{\delta})$ depend on depend on (μ, Σ, μ^*). This provides a closed form solution and can also be used to show that $\text{Var}(R_p)$ is a U-shaped parabola, as a function of μ^*. (See below for a proof.)

Remark 3.6 *(Alternative expression for the portfolio weights) Define the scalars a, b and c as $a = \mu'\Sigma^{-1}\mu, b = \mu'\Sigma^{-1}\mathbf{1}$, and $c = \mathbf{1}'\Sigma^{-1}\mathbf{1}$. Then, calculate the* *scalars (for a given required return μ^*)*

$$\tilde{\lambda} = \frac{c\mu^* - b}{ac - b^2} \text{ and } \tilde{\delta} = \frac{a - b\mu^*}{ac - b^2}$$

to get (3.9). To show this, solve (3.8) and rearrange.

Proof that the MVF is a parabola*. From (3.9), the portfolio variance is

$$\text{Var}(R_p) = w'\Sigma w = (w'\mu\tilde{\lambda} + w'\mathbf{1}\tilde{\delta}) = \mu^*\tilde{\lambda} + \tilde{\delta},$$

where we use the facts that $w'\mu = \mu^*$ and $w'\mathbf{1} = 1$. Use the definitions of $(\tilde{\lambda}, \tilde{\delta})$ in Remark 3.6, complete the square and simplify to get

$$\text{Var}(R_p) = \frac{c(\mu^* - b/c)^2}{ac - b^2} + \frac{1}{c},$$

where (a, b, c) depend on (μ, Σ), not on μ^*. This is a U-shaped parabola when μ^* is on the horizontal axis and $\text{Var}(R_p)$ on the vertical. The minimum is $1/c$, which is the variance of the global minimum variance portfolio (defined below). See Pennacchi (2008) 2 for a detailed discussion. □

Another way to construct the MVF of risky assets is to retrace it by *combining any two portfolios on the frontier*. This is sometimes referred to as the "two-fund theorem", although that should not be confused with the two-fund separation theorem discussed in later chapters. 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 MVF. 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), but we could have used other portfolios. (See below for a proof.)

Proof that w_κ in (3.10) is on the MVF*. Notice that (3.9) can be rewritten as $w = \omega + \phi\mu^*$, where (ω, ϕ) depend on (μ, Σ), but not on μ^*. (The expressions are $\omega = \Sigma^{-1}(\mathbf{1}a - \mu b)/(ac - b^2)$ and $\phi = \Sigma^{-1}(\mu c - \mathbf{1}b)/(ac - b^2)$.) It follows that a portfolio (p) with weights $w_p = \omega + \phi\mu_p^*$ must be somewhere on the MVF. In particular, consider a combination of two MVF portfolios (w_g and w_T, say), that is, $w_p = \kappa w_g + (1-\kappa)w_T$. Notice that $w_p = \kappa(\omega + \phi\mu_g^*) + (1-\kappa)(\omega + \phi\mu_T^*)$, where (μ_g^*, μ_T^*) are the expected returns on the two respective MVF portfolios. Since $\mu_p^* = \kappa\mu_g^* + (1-\kappa)\mu_T^*$, the weights for portfolio p the can be written as

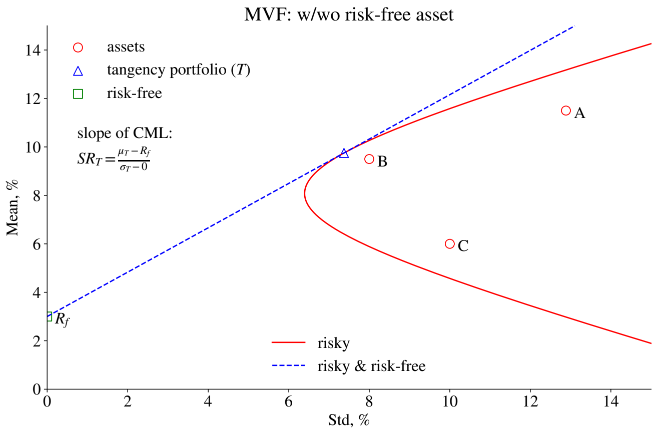


Figure 3.7: Mean-variance frontiers, w/wo risk-free asset. The properties of the investable assets (A, B, and C) are shown in Table 3.1.

$w_p = \omega + \phi\mu_p^*$. Vary μ_p^* (or equivalently, κ) to trace out the frontier. □

3.2 The Mean-Variance Frontier of Risk-Free and Risky Assets

We now add a risk-free asset with return R_f and notice that the restriction that $\mathrm{E}\, R_p = \mu^*$ can be written as

$$w'\mu + (1 - w'\mathbf{1})R_f = w'\mu^e + R_f = \mu^*, \tag{3.11}$$

where μ^e the vector of mean excess returns ($\mu - R_f$). Here we use w to denote the vector of portfolio weights on the risky assets only, with $1 - w'\mathbf{1}$ (that is, $1 - \Sigma_{i=1}^n w_i$) as the weight on the risk-free asset. This means that the requirement that all portfolio weights sum to 1 is automatically satisfied.

The minimization problem (3.1) can now be written

$$\begin{aligned} &\min\nolimits_w \; w'\Sigma w \text{ subject to} \\ &w'\mu^e + R_f = \mu^*. \end{aligned} \tag{3.12}$$

When there are no additional constraints, then we can find an explicit solution. In other cases we need to apply numerical optimization. The weights of the risky assets for a portfolio on the mean-variance frontier, at a given required return μ^*,

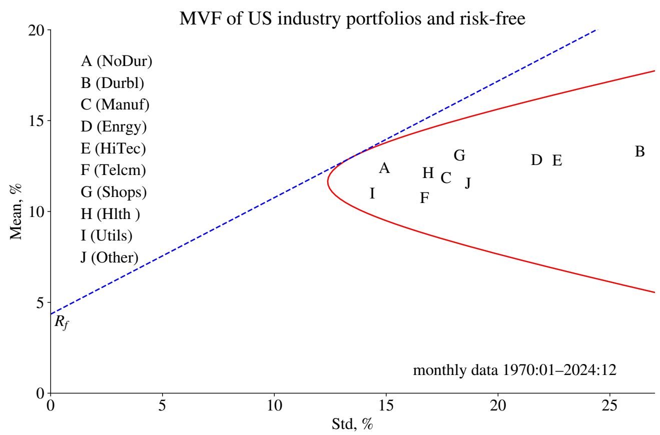


Figure 3.8: Mean-variance frontier from US industry indices

are

$$w = \frac{\mu^* - R_f}{\mu^{e\prime}\Sigma^{-1}\mu^e}\Sigma^{-1}\mu^e. \tag{3.13}$$

As mentioned before, the weight on the risk-free asset is $1 - w'\mathbf{1}$. (See below for a proof.)

Repeating the calculation for different expected return, μ^*, allows us to trace out the entire mean-variance frontier. In the std×mean space, the efficient frontier (the upper part) is just a *line*, called the *Capital Market Line* (CML). See Figure 3.7 for an illustration and Figure 3.8 for an empirical example showing an (ex post) mean-variance frontier from a sample of U.S. data.

Remark 3.7 *(*Alternative way to calculate w) The proof of (3.13) shows that we calculate w by solving the following system of equations*

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

Remark 3.8 *(Minimizing the standard deviation) It can be shown that the solution (3.13) also solves the problem* $\min \operatorname{Std}(R_p)$ *st* $\mathrm{E}\, R_p = \mu^*$ *and* $\Sigma_{i=1}^n w_i = 1$.

Remark 3.9 *(MVF with different lending and borrowing rates*) Figure 3.9 illustrates the MVF when the borrowing rate is higher than the lending rate. The frontier has three segments: (1) a straight CLM between the lending rate and the tangency*

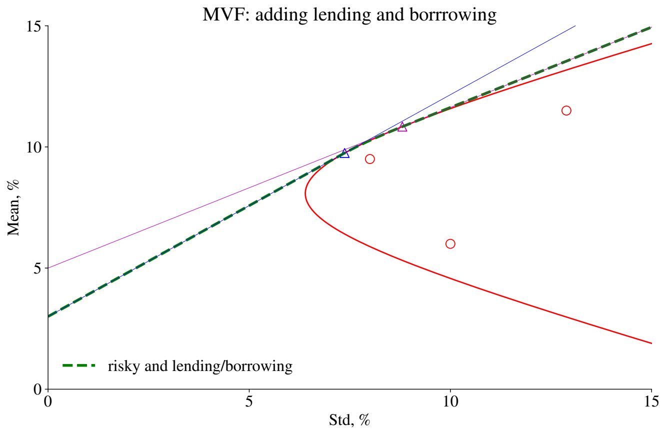


Figure 3.9: Mean-variance frontier with different lending and borrowing rates. The properties of the investable assets (A, B, and C) are shown in Table 3.1.

portfolio defined by that rate; (3) a different straight CLM defined by and starting at the tangency portfolio calculated from the borrowing rate; and (2) a segment (arc) in the middle where the investment is in risky assets only. As the lending and borrowing rates gets closer, this converges to the earlier (single) CLM and a single tangency portfolio.

Proof of (3.13). Define the Lagrangian problem

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

(Variances are denoted σ_{ii} in order to facilitate comparison with the matrix expressions.) The first order condition with respect to the portfolio weights are

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

Similarly, the first order condition with respect to the Lagrange multiplier is

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

Combine as

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

which can be written

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

To simplify further, notice that the first set of equations is $\Sigma w = -\lambda \mu^e$, which can be (partially) solved as $w = -\Sigma^{-1}\lambda\mu^e$. The second set of equations is $\mu^{e\prime} w = \mu^* - R_f$. Use the (partial solution) of w to write this as $-\mu^{e\prime}\Sigma^{-1}\lambda\mu^e = \mu^* - R_f$, which can be solved for as $\lambda = -(\mu^* - R_f)/\mu^{e\prime}\Sigma^{-1}\mu^e$. Finally, using this in the partial solution of w gives (3.13). □

3.3 The Tangency Portfolio

The mean-variance frontier for risky assets only and the frontier for risky assets plus the risk-free asset are tangent at one point—called the *tangency portfolio*: see Figure 3.7. In this case the portfolio weights from (3.8) and (3.13) coincide. Therefore, the portfolio weights of the risky assets (3.13) must sum to unity (so the weight on the risk-free asset is zero) at this value of the required return, μ^*. This gives the portfolio weights of the tangency portfolio

$$w_T = \frac{\Sigma^{-1}\mu^e}{\mathbf{1}'\Sigma^{-1}\mu^e}. \tag{3.14}$$

Proof of (3.14). Put the sum of the portfolio weights in (3.13) equal to one and solve for the μ^* value where that holds. Use in (3.13). □

Notice that Capital Market Line (CML) starts at the location $(\sigma, \mu) = (0, R_f)$ and goes through the point (μ_T, σ_T) where the latter are the mean and standard deviation of the tangency portfolio. It is then clear that the slope of the CML, $(\mu_T - R_f)/(\sigma_T - 0)$, represents the *Sharpe ratio of the tangency portfolio*. The line is thus

$$\mathrm{E}\, R_{opt} = R_f + \sigma_{opt} SR_T. \tag{3.15}$$

Interestingly, the tangency portfolio has the *highest Sharpe ratio of any portfolio* that can be created from the investable assets. See Figure 3.10.

It follows that every portfolio on the CML is a combination of the tangency portfolio and the risk-free asset

$$R_{opt} = v R_T + (1 - v) R_f \tag{3.16}$$

where R_T is the return on the tangency portfolio. Again, see Figure 3.10.

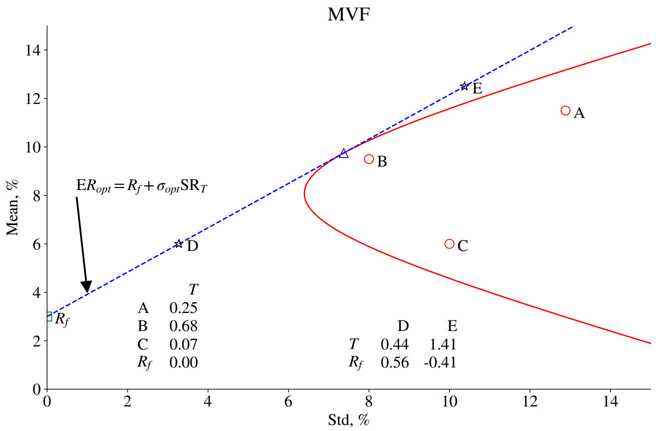


Figure 3.10: Mean-variance frontiers, creating portfolios by combining the tangency portfolio and the risk-free. The properties of the investable assets (A, B, and C) are shown in Table 3.1.

Remark 3.10 *(*Maximising the Sharpe ratio directly.) Maximizing $v'\mu^e/\sqrt{v'\Sigma v}$ gives the following n first order conditions*

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

Setting v equal to the tangency portfolio in (3.14) satisfy those first order conditions. (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 3.11 *(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 (3.14) is then

$$\begin{bmatrix} w_{T,1} \\ w_{T,2} \end{bmatrix} = \frac{1}{\sigma_2^2 \mu_1^e + \sigma_1^2 \mu_2^e} \begin{bmatrix} \sigma_2^2 \mu_1^e \\ \sigma_1^2 \mu_2^e \end{bmatrix}. \tag{3.17}$$

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. (You may notice that scaling the mean returns and/or the variance-covariance matrix does not matter.)

Example 3.12 *(Tangency portfolio, numerical) When* $(\mu_1^e, \mu_2^e) = (8\%, 5\%)$*, the correlation is zero, and* $(\sigma_1^2, \sigma_2^2) = (256\ bp, 144bp)$*, then (3.17) 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 ($\sigma_1 = \sigma_2 = 1$ as a normalization, $\sigma_{12} = \rho$). Then (3.17) becomes

$$\begin{bmatrix} w_{T,1} \\ w_{T,2} \end{bmatrix} = \frac{1}{(\mu_1^e + \mu_2^e)(1 - \rho)} \begin{bmatrix} \mu_1^e - \rho \mu_2^e \\ \mu_2^e - \rho \mu_1^e \end{bmatrix}. \tag{3.18}$$

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 3.13 *(Tangency portfolio, numerical) In the case of (3.18) with* $(\mu_1^e, \mu_2^e) = (8\%, 5\%)$*, 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*

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

Chapter 4

The Inputs to MV Calculations

Mean-variance (MV) analysis and portfolio choice depend on assumptions regarding average returns and the variance-covariance matrix of the investable assets. This chapter discusses how estimates from historical data can help in formulating such assumptions, although judgemental adjustments are likely to be made.

4.1 Introduction

The moments (means, variances and covariances) are *conditional* on the information available at the time of portfolio formation, depend on the *investment horizon* and they may *change over time*.

The *conditional* nature of the moments used in portfolio formation distinguish them different from traditional sample estimates. To illustrate, consider the definition

$$R_{t+1} = \mathrm{E}_t\, R_{t+1} + \varepsilon_{t+1}, \tag{4.1}$$

where R_{t+1} is the return over the investment horizon, $\mathrm{E}_t\, R_{t+1}$ the forecast based on information when then portfolio is formed in period t, and ε_{t+1} is the unforecasted part of the return (news, surprise).

A traditional sample estimate of the variance measures the historical variance in R_t. In contrast, the portfolio formation is based on the variance of the forecast error, ε_t. For most assets, returns are difficult to forecast; hence, the difference between the two measures of variance is small. For instance, if the forecasting model has a coefficient of determination ("R^2") of 0.05, which seems to be close to the upper limits of most return forecasting models, then the variance of ε_t is 0.95 times the variance of R_t. In this situation, the sample variance of R_t might be a good approximation. In contrast, for the risk-free rate, the conditional variance is zero, while sample variance is not (albeit small).

It is also important to consider *time-variation* in the moments. In particular, variances and covariances have considerable (predictable) movements, which motivates using some kind of time-series method for estimation. In addition, *sample estimates can be noisy*, especially when the sample is small, which may motivate forming a compromise between the sample estimates and a priori information ("shrinkage").

4.2 The Market Model: Betas

The beta (slope coefficient) from the *market model* is often used to describe the cyclicality of an asset. It is useful as a statistical description of the returns. The regression is

$$\begin{aligned} R^e_{it} &= \alpha_i + \beta_i R^e_{mt} + \varepsilon_{it}, \text{ where} \\ \mathrm{E}\,\varepsilon_{it} &= 0,\ \mathrm{Cov}(\varepsilon_{it}, R^e_{mt}) = 0. \end{aligned} \tag{4.2}$$

Here, R^e_{it} is the excess return on asset i in period t, while R^e_{mt} is the market excess return in the same period. The regression is done on time series (R^e_{it} and R^e_{mt} for $t = 1, 2, ..., T$). As usual, the regression slope is $\beta_i = \sigma_{im}/\sigma^2_m$. This regression may use the excess returns as indicated above, or the net returns. The two assumptions (the residual has a zero mean and is uncorrelated with the regressor) are standard in regression analysis. See Figures 4.1 for an illustration.

Empirical Example 4.1 *See Figures 4.2–4.3 for results on U.S. industry portfolios. The β value are scattered around 1 and the R^2 values are high (around 0.65 on average). See also Table 4.1 for some alternative assets. In general, we need to move away from equities to get β values close to zero (or negative).*

The beta of a portfolio is the portfolio of betas. That is, for a portfolio of assets, the beta is

$$\beta_p = \textstyle\sum_{i=1}^n w_i \beta_i. \tag{4.3}$$

(This follows from the fact that $\mathrm{Cov}(w_i R_i + w_j R_j, R_m) = w_i \sigma_{im} + w_j \sigma_{jm}$. Dividing by σ^2_m gives the betas.)

We will later discuss how the market model can help in estimating the variance-covariance matrix of the assets. It is also clear that the β values can be useful in *portfolio formation*. For instance, suppose we want to combine assets 1 and 2 in

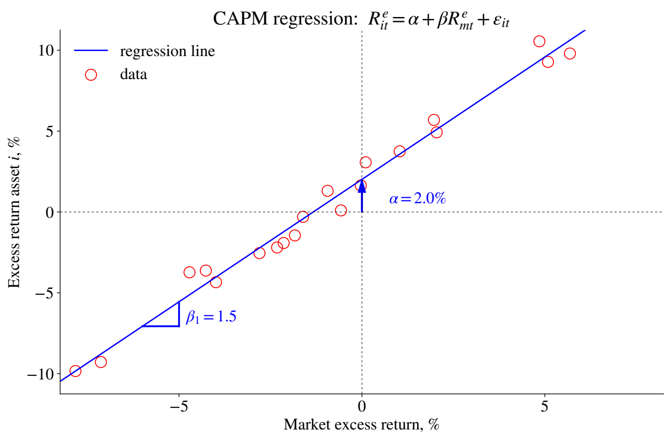


Figure 4.1: CAPM regression

such a way that our overall position is market neutral ($\beta_p = 0$). This can be done by choosing the portfolio weights $(w_1, w_2) = (1, -\beta_1/\beta_2)$, combined with a position in the risk-free rate. Also, if the investor wants a portfolio with a particular beta (β_q), then this can be achieved by investing β_q in the market portfolio and $1 - \beta_q$ in the risk-free asset.

The result in (4.3) also shows that the (value weighted) β of all assets must equal 1. (This follows from the fact that the value weighted portfolio of all assets equals the market portfolio—and regressing the market on itself must give a slope of 1.)

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

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

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 net return on holding asset i between $t-1$ and t; and (b) w_{it} is the market capitalization of asset i relative to the total market capitalization of all n assets, measured at the end of period $t-1$. Most of the important indices are of this sort. Instead, if R_{it} only includes the capital gain of holding asset i, then the index is a price index. In other cases, the weights may 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).

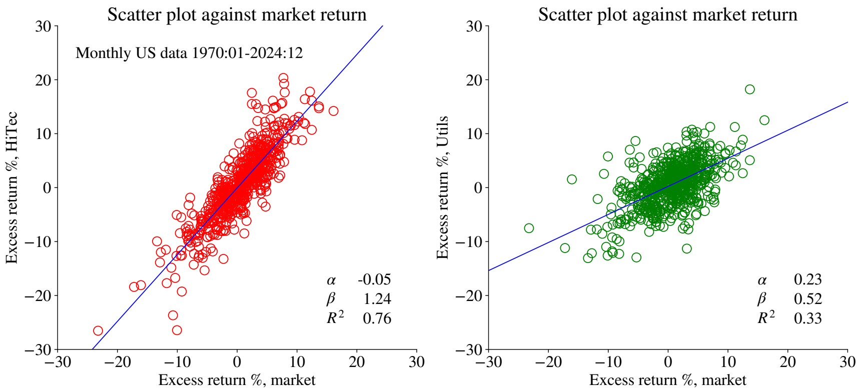


Figure 4.2: Scatter plot against market return

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 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).*

	β	R^2
MSCI world	0.93	0.92
CT hedge funds	0.26	0.40
Global govt bonds	0.07	0.02
Gold	0.02	0.00
Oil	0.26	0.02

Table 4.1: β and R^2 against the U.S. equity market, monthly returns, 1993:12-2024:12.

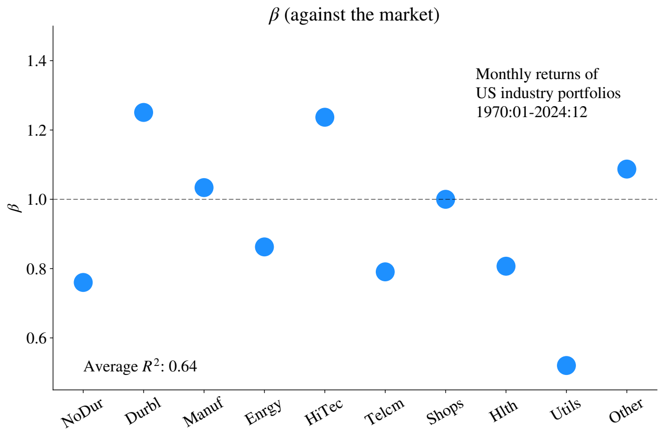


Figure 4.3: βs of US industry portfolios

4.2.1 Estimating Historical Beta: OLS and Other Approaches

It is sometimes argued that the OLS estimate of beta on a historical sample may not be the best forecast of the beta for a future time periods (see, for instance, Blume (1971)). As a potential solution, we could apply a shrinkage towards the average beta (which is 1)

$$\beta = \eta\hat{\beta}_{OLS} + (1 - \eta)1. \tag{4.4}$$

This could be motivated by empirical findings (Blume (1975)) or by a Bayesian principle (see Greene (2018) 16). In the latter case, η would be higher if the sample is long and when the fit is good.

To capture time-variation in betas, we could either estimate on a moving data window or apply an exponentially weighted moving average estimate (EWMA). The latter is a weighted OLS where an observation s periods ago gets the weight λ^s where λ is close to one (for instance, 0.95).

Empirical Example 4.5 *See Table 4.2 for an evaluation of several methods: most are of the form (4.4), but an EWMA approach is also considered. A negative number indicate that the method is better than OLS.*

	OLS adj $0.67\hat{b}+0.33$	OLS adj $0.5\hat{b}+0.5$	OLS adj $0.33\hat{b}+0.67$	1 1	EWMA $0.5\hat{b}+0.5$
$\lvert$error in $\beta\rvert$	-6.1	-5.9	-2.9	8.7	-10.2

Table 4.2: Absolute forecast errors of future betas, as a percentage difference to OLS: the average $|$next 2 year β − predicted $\beta|$ compared to the results from OLS. A negative number is better performance than OLS. The models are estimated on moving 10-year windows and EWMA uses $\lambda = 0.95$. 25 FF portfolios, monthly data for 1970:01-2024:12.

4.2.2 Fundamental Betas

Another way to improve the beta forecasts is to incorporate 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 (see Rosenberg and Guy (1976) for an early paper and Damodaran (2012) for a more recent text) 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.

4.3 Estimation of the Covariance Matrix of the Asset Returns

There are several issues with estimating variance-covariance matrices: *(1)* the number of parameters increase very quickly as the number of assets increases ($n(n+1)/2$ with n assets, for instance 5,050 for 100 assets); *(2)* there may not be relevant historical data; *(3)* historical estimates have proven somewhat unreliable for future periods due to small sample issues and time-variation of the parameters.

Remark 4.6 *(Fama-French portfolios) The 25 FFF portfolios (used in the examples below) 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 4.3 the portfolio in row 2, column 3 (portfolio 8) belong to the 20%-40% largest firms and the 40%-60% firms with the highest book value/market value.*

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

Table 4.3: Numbering of the FF portfolios.

4.4 Covariance Matrix with Time-Varying Parameters

To handle the time-variation, we could apply a simple EWMA estimator (cf. the RiskMetrics approach of JP Morgan (1996))

$$\mu_{it} = \lambda \mu_{i,t-1} + (1-\lambda) x_{i,t-1}, \tag{4.5}$$

$$\sigma_{ij,t} = \lambda \sigma_{ij,t-1} + (1-\lambda)(x_{i,t-1} - \mu_{i,t-1})(x_{j,t-1} - \mu_{j,t-1}), \tag{4.6}$$

with $0 \leq \lambda \leq 1$ and where x_{it} is element i of the vector x_t. The first equation provides a time-varying estimate of the mean and the second equation of the covariance between x_{it} and x_{jt} (set $i = j$) for the variance. In many application on daily data a value of $\lambda \approx 0.94$ is common. For monthly data, often slightly lower values.

4.5 Covariance Matrix with Average Correlations

A commonly used method to address the instability of the variance-covariance matrix, caused by excessive parameters or insufficient data, is to replace the historical correlation with an average historical correlation. To do that, estimate ρ_{ij} on historical data, but use the average estimate $\bar{\rho}$ as the "forecast" of all correlations and calculate adjusted covariances as

$$\sigma_{ij} = \bar{\rho} \sigma_i \sigma_j \text{ for } i \neq j. \tag{4.7}$$

(The variances are not adjusted.) Notice that $\bar{\rho}$ is the average of the $n(n-1)/2$ elements below (or above) the main diagonal of the correlation matrix.

4.6 Covariance Matrix from a Single-Index Model

The single-index model is another way to reduce 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 (4.2)

$$\text{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0, \tag{4.8}$$

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_{mt}).

If (4.2) and (4.8) are true, then the covariance of assets i and j is

$$\sigma_{ij} = \beta_i \beta_j \sigma_{mm} \text{ for } i \neq j, \tag{4.9}$$

where σ_{mm} is the variance of R_m. For instance, when there is too little (relevant) historical data to estimate reliable covariances, then we could first assess the betas of the assets (for instance, based on firm characteristics) and then draw conclusions about the covariances. The variances are often estimated directly from the sample. See Elton, Gruber, Brown, and Goetzmann (2014) 7–8 for more details on index models.

Example 4.7 *(Two assets) Let* $[\beta_1, \beta_2] = [0.9, 1.1]$, $[\text{Var}(\varepsilon_{1t}), \text{Var}(\varepsilon_{2t})] = [100, 25]/100^2$, *and* $\sigma_{mm} = 225/100^2$. *Then*

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

Proof of (4.9). By using (4.2) and (4.8) and recalling that $\text{Cov}(R_m, \varepsilon_i) = 0$ direct calculations give that the covariance of assets i and j ($i \neq j$) is (recalling also that $\text{Cov}\left(\varepsilon_i, \varepsilon_j\right) = 0$)

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

□

4.7 Covariance Matrix from a Multi-Index Model

The multi-index model is an extension of the single-index model

$$R^e_{it} = a_i + b_i' I_t + \varepsilon_{it}, \text{ where} \tag{4.10}$$
$$\mathrm{E}\,\varepsilon_{it} = 0, \ \mathrm{Cov}(\varepsilon_{it}, I_t) = \mathbf{0}, \text{ and } \mathrm{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0,$$

where I_t is a vector of indices. 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 study whether the residuals are approximately uncorrelated. If not, then adding a second index might improve the model.

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

$$\sigma_{ij} = b_i' \Omega b_j \text{ for } i \neq j. \tag{4.11}$$

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).

4.8 Covariance Matrix From A Shrinkage Estimator

The historical sample covariance matrix, S, can exhibit significant noise in small samples. One way of handling that is to "shrink" the sample covariance matrix towards a target, F, as

$$\Sigma = \delta F + (1 - \delta) S, \text{ where } 0 \leq \delta \leq 1. \tag{4.12}$$

Ledoit and Wolf (2003) suggest an F matrix from the single index model and Ledoit and Wolf (2004) instead suggest an F matrix which implies the same correlations of all assets. See the previous sections for how to construct such F matrices. In both cases, the diagonal elements of F are the same as in S. The articles develop algorithms for calculating an approximately optimal value of δ, which tend to be large in small samples and with crude targets.

Empirical Example 4.8 *Table 4.4 suggest that with the 25 FF assets δ is small for* *long samples, but may be non-trivial for shorter samples. With more assets, the δ value is likely to be larger. An alternative approach to choose δ is to investigate the performance on earlier samples.*

	Full sample	10-year samples
constant correlation	0.11	0.64
single index model	0.02	0.11

Table 4.4: Shrinkage parameter δ in Ledoit and Wolf's (2003,2004) covariance estimator $\delta F + (1-\delta) S$ The target matrix F is either a constant correlation covariance matrix or the covariance matrix from a single index model. 25 FF portfolios, monthly data for 1970:01–2024:12. The result for the 10-year samples is the average across moving 10-year data windows.

4.9 An Evaluation of Different Approaches

This section presents a simple assessment the different methods. However, it does not make any general claims as the conclusions depend on *(a)* which assets; *(b)* sample period and time horizons; and *(c)* various modelling parameters.

Empirical Example 4.9 *Table 4.5 present results for the 25 FF portfolios. Several models are estimated on moving 10-year data windows (except the exponentially weighted moving average, EWMA, method which follows (4.5)–(4.6)). We focus on the correlations, since most methods share the same approach for the standard deviations. For each method and period, the implied correlations are calculated and then compared with the actual (realised) correlations for two years after the estimation window. Then, the data window is moved one month and the procedure is repeated. The table show the average (across time and assets) of absolute forecast error of the correlation.*

4.10 Estimating Expected Returns

A later chapter will discuss return predictability at some length. For now it suffices to say that even the best prediction models have limited performance, often with coefficient of determination ("R^2") below 5%. In particular, it turns out that it is hard to beat the historical average return as a predictor.

average corr	1-factor model	3-factor model	shrink to avg corr	shrink to 1-factor	EWMA $\lambda = 0.95$
12.4	40.5	2.4	6.5	1.2	-5.0

Table 4.5: Average absolute forecast errors of future correlations, as a percentage difference to the sample correlation: average |correlation next 2 years - predicted correlation| compared to the sample correlation. A negative number is better performance than the sample correlation. All models (except EWMA) are estimated on moving 10-year windows. The shrinkage approach reports results from covariance matrix $= \delta F + (1-\delta)S$ with δ optimally chosen as in Ledoit and Wolf (2003,2004). 25 FF portfolios, monthly data for 1970:01-2024:12.

Chapter 5

Portfolio Choice

This chapter finds the *optimal portfolio* for an investor who cares about average returns and volatility, not other moments. This turns out to be a particular portfolio on the mean-variance frontier. When the investable assets include both risky and risk-free asset, then it is a mix of the tangency portfolio and the risk-free asset.

5.1 Portfolio Choice with MV Preferences

This chapter discusses optimal portfolio choice when the investor has mean-variance (MV) preferences and can invest in both risky assets and a risk-free asset.

As in earlier chapters, the beliefs about the returns (in particular, their average returns μ and variance-covariance matrix Σ) are taken for granted. The analysis is focused on optimal portfolio choice, given those beliefs about the investable assets.

The investor chooses the portfolio weights to maximize expected utility

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

$$R_p = v'R + (1 - w'\mathbf{1})R_f = v'R^e + R_f. \tag{5.2}$$

The k parameter indicates the degree of risk aversion. (Dividing k by 2 is made for convenience: it makes the equation for the optimal portfolio choice look a bit less involved.) The portfolio return in (5.2) assumes investment in risky assets (vector of portfolio weights v) and a risk-free asset (portfolio weight $1 - w'\mathbf{1}$). Notice that this expression automatically imposes the restriction that all portfolio weights sum to one. As before, the expectation and the variance summarise the beliefs of the investor, conditional on the information available at the time of the investment.

	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 three assets in some 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, $\times 100^2$).

The optimisation problem is illustrated in Figure 5.1. This figure shows utility contours which are combinations of $\mathrm{E}\, R_p$ and $\mathrm{Std}(R_p)$ so that expected utility, $\mathrm{E}\, U(R_p)$, in (5.1) is constant. Contours further to the upper left have higher expected utility and are thus preferred. In contrast, the capital market line (CML) shows what is possible to achieve: points on or below the line.

The optimization problem in (5.1) is to move to a point *as far to the upper left as possible*. Clearly, this will be a point on the CML, so the optimal portfolio is a mix of the risk-free and the tangency portfolio, which will later be referred to as the "two-fund separation theorem".

Also, see Figure 5.2 for an illustration of how the risk aversion determines *which point on the CML* that is optimal: lower risk aversion (k) will lead the investor to accept more risk in exchange for a higher average return.

Remark 5.1 *(Utility contours*) Let u be a fixed level of expected utility and rewrite (5.1) as $u + \frac{k}{2}\,\mathrm{Std}(R_p)^2 = \mathrm{E}\, R_p$. For a given value of $\mathrm{Std}(R_p)$ this gives the required $\mathrm{E}\, R_p$ needed to get expected utility u.*

5.2 A Single Risky Asset and a Risk-Free Asset

Suppose initially that there is a single investable risky assset. The investor then maximizes (5.1) but where (5.2) simplifies to

$$R_p = vR^e + R_f. \tag{5.3}$$

Remark 5.2 *(Real or nominal returns) The objective function (5.1) makes more sense if the returns are real, that is, nominal returns minus inflation. During periods (or over horizons) when inflation is fairly stable, the practical difference is small.*

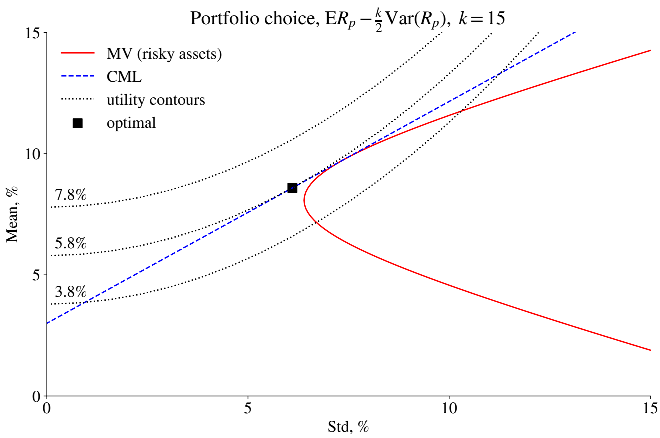


Figure 5.1: Iso-utility curves, mean-variance utility. The calculations use the properties of the assets in Table 5.1.

However, it might matter in other periods. In particular, the risk profile of an asset will then depend on how its return covaries with inflation. For instance, equity returns are often considered to be better hedges against inflation than bond returns.

Use the budget constraint in the objective function to get

$$\begin{aligned}\mathrm{E}\,U(R_p) &= \mathrm{E}(vR^e + R_f) - \frac{k}{2}\mathrm{Var}(vR^e + R_f) \\ &= v\mu^e + R_f - \frac{k}{2}v^2\sigma^2,\end{aligned} \tag{5.4}$$

where (μ^e, σ^2) denote the investor's beliefs about the mean excess return and variance of the risky asset. In the second equation we use the fact that R_f is known.

The first order condition for an optimum ($d\,\mathrm{E}\,U(R_p)/dv = 0$) is

$$\mu^e - kv\sigma^2 = 0, \tag{5.5}$$

which trades off how a marginal increase of v gives a higher expected return but also volatility.

Solve for the optimal portfolio weight of the risky asset as

$$v = \frac{1}{k}\frac{\mu^e}{\sigma^2}. \tag{5.6}$$

The weight on the risky asset is increasing in the expected excess return of the

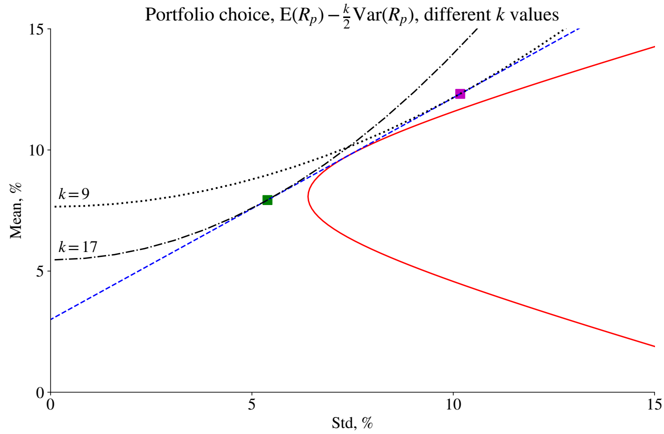


Figure 5.2: Iso-utility curves, mean-variance utility (different risk aversions). The calculations use the properties of the assets in Table 5.1.

risky asset, but decreasing in the risk aversion and variance. See Figure 5.3, which also illustrates that the objective function is concave, meaning that the first order condition is both necessary and sufficient. (From (5.5) we also see that the 2nd-order derivative is $-k\sigma^2$, which is negative.)

Example 5.3 *(Portfolio choice) If* $\mu^e = 6.5\%$, $\sigma_i = 8\%$ *and* $k = 25$, *then* $v \approx 0.41$. *Instead, with* $k = 10$, $v \approx 1.02$.

Remark 5.4 *(*Why not use* $\mathrm{E}\, R_p - k\,\mathrm{Std}(R_p)$*?) Because it may not have a finite optimum as the objective function is not strictly concave. Too see this, consider changing (5.4) to* $v\mu^e - k\sqrt{v^2\sigma^2}$ *and suppose* $v \geq 0$ *is optimal. The objective function is then* $v(\mu^e - k\sigma)$ *where* $v = \infty$ *is optimal if* $\mu^e > k\sigma$. *The problem is that both the average returns and the standard deviation are linear in* v. *Instead, if we were to maximize* $v\mu^e - k(v^2\sigma^2)^{0.51}$ *(notice the* 0.51 *instead of* 0.5*), then the problem is well behaved.*

5.3 Several Risky Assets and a Risk-Free Asset

We now consider the case with n investable risky assets and a risk-free asset. Combining (5.1) and (5.2) gives

$$\mathrm{E}\, U(R_p) = v'\mu^e + R_f - \frac{k}{2} v'\Sigma v, \tag{5.7}$$

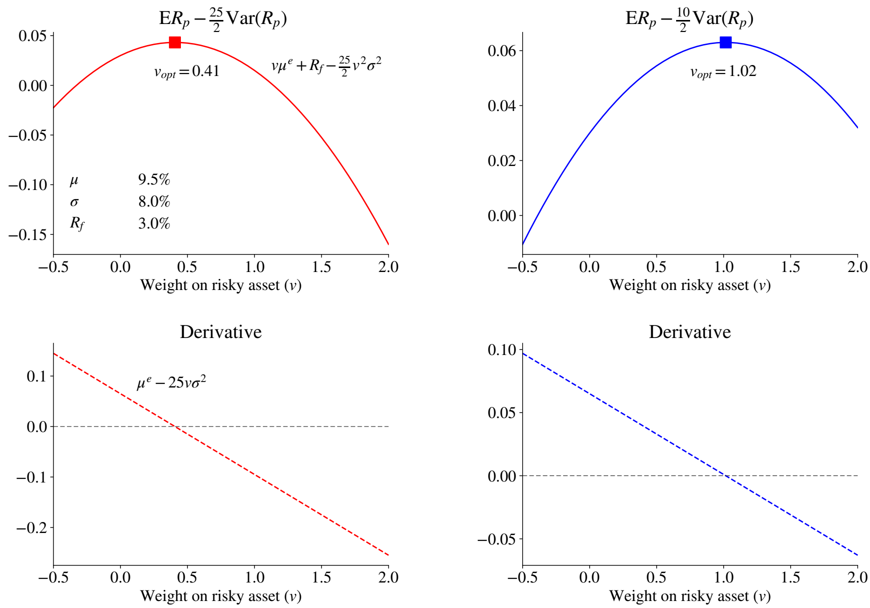


Figure 5.3: Portfolio choice, a single risky and a risk-free asset

where μ^e the n-vector of average excess returns and Σ is the $n \times n$ covariance matrix of the returns. As before, these moments represent the beliefs of the investor, conditional on the information available at the time of the investment.

The first order conditions (for the vector v) are that the partial derivatives with respect to v are zero

$$\mu^e - k\Sigma v = \mathbf{0}, \tag{5.8}$$

which can be solved as

$$v = \Sigma^{-1}\mu^e/k. \tag{5.9}$$

Notice that the weight on the risk-free asset is $1 - v'\mathbf{1}$, where $\mathbf{1}$ is a (column) vector of ones. We will later provide an interpretation of the first order conditions.

Remark 5.5 *For two assets, (5.9) can be written*

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

where we use σ_{ii} to indicate the variance of asset i, since this facilitates the com-

	μ, %	Σ, bp	
		A	B
A	11.5	166	34
B	9.5	34	64

Table 5.2: Characteristics of the two assets in some 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, $\times 100^2$). The risk-free rate is 3%.

parison with the matrix expressions. Notice that the denominator $\sigma_{11}\sigma_{22} - \sigma_{12}^2$ is positive since the correlation $\sigma_{12}/(\sigma_1\sigma_2)$ is between -1 and 1. This means that

$$v_i > 0 \text{ if } SR_i > \rho SR_j,$$

where ρ is the correlation. 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.

Example 5.6 *((5.9) with two assets) Let $\Sigma = \begin{bmatrix} 166 & 34 \\ 34 & 64 \end{bmatrix} / 100^2$, $\mu^e = \begin{bmatrix} 5.5 \\ 3.5 \end{bmatrix} / 100$ and $k = 9$. Then*

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \approx \begin{bmatrix} 67.6 & -35.9 \\ -35.9 & 175.3 \end{bmatrix} \begin{bmatrix} 0.055 \\ 0.035 \end{bmatrix} \frac{1}{9} \approx \begin{bmatrix} 0.27 \\ 0.46 \end{bmatrix}.$$

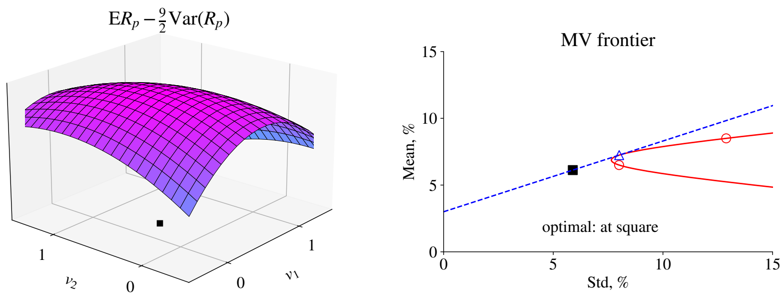
The weight on the risk-free is (approximately) $1 - 0.27 - 0.46 = 0.27$. See also Figure 5.4.

Remark 5.7 *(Covariance of portfolios) Σv is the vector of covariances of each asset return, R, with the return on the portfolio $v'R$. Also, $\partial v'\Sigma v/\partial v = 2\Sigma v$, is the marginal contribution of each of the assets to the variance of the portfolio.*

Remark 5.7 says that Σv is the n-vector of covariances of each investable asset with the optimal portfolio. This means that row i of the first order conditions (5.8) can be written

$$\mu_i^e - k\,\mathrm{Cov}(R_i, R_v) = 0, \tag{5.10}$$

The first term is the marginal increase in the portfolio excess return from increasing the weight on asset i slightly—and financing it by borrowing. The second term is the risk aversion k times the marginal increase in portfolio variance (divided by 2).



	optimal	tangency	ratio
asset 1	0.27	0.37	0.74
asset 2	0.46	0.63	0.74
risk-free	0.26	0.00	

Figure 5.4: Choice of portfolios weights. The calculations use the properties of the assets in Table 5.2.

At at the optimum, the two terms (the "benefit" and the "cost" of increasing the position in asset i) are equal. See Figure 5.5 for an illustration. Off the optimum, one is larger than the other—so it is beneficial to change the portfolio. Increasing v_i does not change the first term of (5.10) which is constant at μ_i^e but it will change the second term. The reason for the latter is that a higher v_i value will make the portfolio return more similar to R_i and thus increase the covariance.

Proof of (5.9). With $n = 2$ the portfolio return can be written $R_p = v_1 R_1^e + v_2 R_2^e + R_f$,so the objective is

$$\mathrm{E}\,U(R_p) = v_1\mu_1^e + v_2\mu_2^e + R_f - \frac{k}{2}(v_1^2\sigma_{11} + v_2^2\sigma_{22} + 2v_1v_2\sigma_{12}),$$

where σ_{ii} denotes the variance of asset i and σ_{12} the covariance of asset 1 and 2. The first order conditions are

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

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

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

This is the same as (5.8). □

Remark 5.8 *(Several risky assets, no risk-free*) 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*)* $\mu - k\Sigma v - \theta\mathbf{1} =$

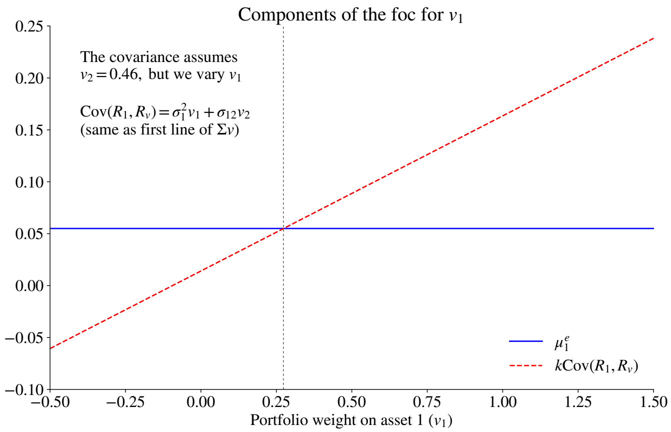


Figure 5.5: Choice of portfolios weights, first order condition. The calculations use the properties of the assets in Table 5.2.

0. *Rewrite and use the restriction* ($\mathbf{1}'v=1$) *to write on matrix form*

$$\begin{bmatrix} k\Sigma & \mathbf{1} \\ \mathbf{1}' & 0 \end{bmatrix} \begin{bmatrix} v \\ \theta \end{bmatrix} = \begin{bmatrix} \mu \\ 1 \end{bmatrix}.$$

Solve for $[v; \theta]$.

5.4 MV Preferences Gives a Portfolio on the MV Frontier

It is evident that the optimal portfolio (5.9) is a scaling up/down of the *tangency portfolio* (see previous chapters)

$$w_T = \frac{\Sigma^{-1}\mu^e}{\mathbf{1}'\Sigma^{-1}\mu^e}. \tag{5.11}$$

This confirms the result previously discussed in Section 5.1 and illustrated in Figure 5.1.

To be precise, the optimal portfolio can be written

$$v = cw_T, \text{ where } c = \mathbf{1}'\Sigma^{-1}\mu^e/k \tag{5.12}$$

and $1 - c$ in the risk-free asset. The return on any optimal portfolio is thus a mix of

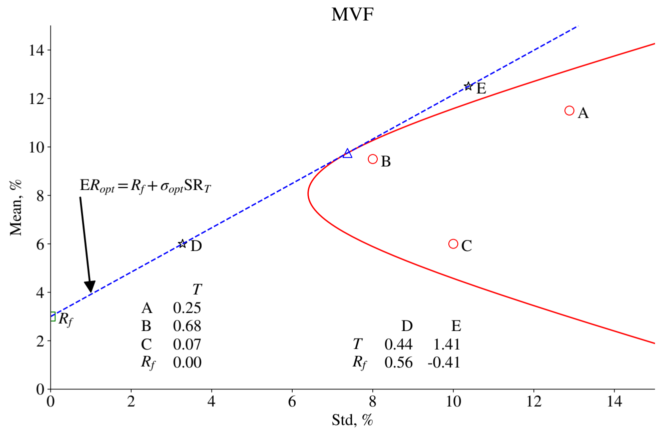


Figure 5.6: Mean-variance frontiers. The properties of the investable assets (A, B, and C) are shown in Table 5.1.

the risk-free and tangency portfolio returns

$$\begin{aligned} R_{opt} &= v' R^e + R_f \\ &= c R^e_T + R_f . \end{aligned} \tag{5.13}$$

This means that the optimal portfolio is on the CML and can be constructed by combining the tangency portfolio and the risk-free asset. This result is often called the "two-fund separation theorem". This has important practical consequences, since it suggests that only two "funds" (one mimicking the tangengcy portfolio, the other being a risk-free asset) are needed to form optimal portfolios. See Figure 5.6 for an illustration.

Equation (5.13) also shows that the beta of the optimal portfolio is

$$\beta_{opt} = c. \tag{5.14}$$

(This follows directly from the fact that regressing $cR^e_T + R_f$ on R^e_T must give a slope of c, since R_f is constant.) Equation (5.12) then shows that risk averse investors (high k) will chose portfolios with low β and vice versa.

Example 5.9 *(Portfolio choice to get a desired β) To construct a portfolio with $\beta = 1.2$ against the tangency portfolio, invest $c = 1.2$ in the tangency portfolio and -0.2 in the risk-free.*

Remark 5.10 *(The mathematics of why* $\max \mathrm{E}\, R_p - k\,\mathrm{Var}(R_p)/2$ *gives a MV portfolio*)* *The efficient set solves the problem* $\max \mathrm{E}\, R_p$ *subject to* $\mathrm{Var}(R_p) \leq q$ *(where we vary* q *to trace out the efficient set). Notice that maximizing* $\mathrm{E}\, R_p - k\,\mathrm{Var}(R_p)/2$ *can be thought of as the Lagrangian formulation of the efficient set problem.*

Chapter 6

CAPM

This chapter starts from the earlier portfolio choice (asset demand) results, and then forms a market equilibrium to extract implication for the pricing of different assets. This leads to the capital asset pricing model (CAPM), where the average return of an asset is explained by its covariance (or β) with the market return. The last section tests this empirically, with mixed results.

6.1 Beta Representation of Expected Returns

6.1.1 Beta Representation: Definition

The beta representation (and eventually also CAPM as developed by Sharpe (1964), Lintner (1965) and Mossin (1966)) follows from the analysis of portfolio choice based on mean-variance preferences.

From an earlier chapter, we notice two things. First, the *first order conditions* for optimal portfolio choice are the n equations in

$$\mathrm{E}\, R^e = k\Sigma v. \tag{6.1}$$

Recall that Σv is a vector of covariances of each asset with the v-portfolio. Second, the optimal portfolio weights (v) are *proportional to the tangency portfolio*, w_T,

$$v = c w_T, \tag{6.2}$$

where c is a scalar. Notice that these expressens depend on beliefs about the average returns (μ) and their variance-covariance matrix (Σ) of the investable assets. All the results in this section are therefore dependent on those beliefs.

Combining these two observations shows that the first order conditions for optimal portfolio choice can be written

$$\mathrm{E}\,R^e = kc\Sigma w_T, \text{ or} \tag{6.3}$$

$$\mathrm{E}\,R^e_i = kc\sigma_{iT}, \text{ for } i = 1, \dots, n, \tag{6.4}$$

where k is the risk aversion and σ_{iT} is shorthand notation for the covariance of R_i and R_T, which is the same as element i of Σw_T. I use the notation $\mathrm{E}\,R^e_i$ for the left hand side (rather than the equivalent μ^e_i) to suggest that this is a result, not "data."

We can express this as a *beta representation*. Let μ^e_T be the expected excess return on the tangency portfolio and rewrite (6.4) as

$$\mathrm{E}\,R^e_i = \beta_i \mu^e_T \text{ where } \beta_i = \sigma_{iT}/\sigma^2_T, \text{ or} \tag{6.5}$$

$$\mathrm{E}\,R_i = R_f + \beta_i \mu^e_T. \tag{6.6}$$

(See below for a proof). Plotting $\mathrm{E}\,R^e_i$ or $\mathrm{E}\,R_i$ against β_i gives the *security market line*, see Figure 6.1.

It is important to acknowledge that this expression does not say anything about *causality*: it just shows how expected returns and betas relate to each other according to the first order conditions for optimal portfolio choice.

Proof of (6.5). Premultiply both sides of (6.3) by w'_T to get $w'_T\mu^e = kcw'_T\Sigma w_T$ which is the same as $\mu^e_T = kc\sigma^2_T$. Solve for kc and use in (6.4). The end of the section presents an alternative proof. $\square$

The β_i is clearly the slope coefficient in a (time series) OLS regression

$$R^e_{it} = \alpha_i + \beta_i R^e_{Tt} + \varepsilon_{it}, \tag{6.7}$$

where R^e_{it} is the excess return on asset i in period t.

Remark 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 matrix Σ of the investable assets, then we can also use the fact that $\beta_i = \sigma_{iT}/\sigma^2_T$ where $\sigma_{iT} = w'_i\Sigma w_T$. (Here, w_i is trivial: 1 in position i and 0 elsewhere.) Using the asset price characteristics in Table (6.1), together with the weights of the tangency portfolio gives the β values in Figure 6.1.*

Example 6.2 *(β_i vs $\mathrm{E}\,R_i$) With $R_f = 3\%$ and $\mu_T = 9.75\%$ (so $\mu^e_T = 6.75\%$) we*

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

Table 6.1: Characteristics of the three assets in some 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, $\times 100^2$).

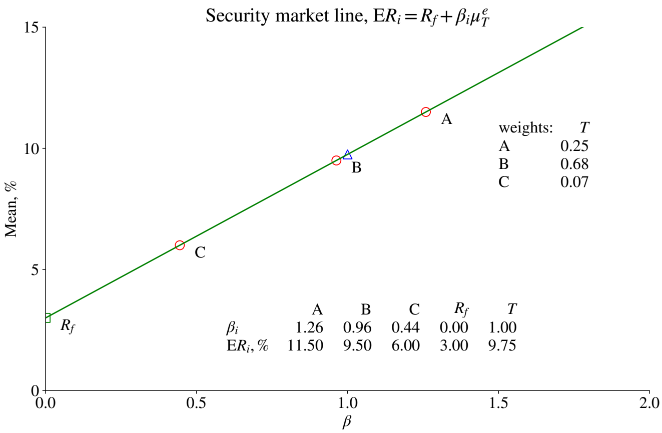


Figure 6.1: Security market line. The properties of the investable assets (A, B, and C) are shown in Table 6.1.

get

β_i	E R_i	*Comment*
0.44	6.0%	*low β, low avg. return*
1.5	13.12%	*high β, high avg. return*
1	9.75%	*same risk as market*
0	3%	*no risk*
−0.5	−0.38%	*the opposite of risk*

Proof of (6.5), alternative*. The foc are $\Sigma v^z = \mathrm{E}\, R^e / k^z$, where k^z is the risk aversion and v^z is the vector of portfolio weights of investor z. Beliefs are assumed to be shared by all investors. Assume all investors have the same capital (or that wealth is uncorrelated with risk aversion) and average the foc across the Z investors,

$\Sigma_{z=1}^{Z}\Sigma v^z/Z$, to get $\Sigma v_m = \mathrm{E}\, R^e/k^*$, where v_m is the market (average) portfolio and where $k^* = 1/\Sigma_{z=1}^{Z} 1/(k^z Z)$ defines an aggregate risk aversion k^*. Row i of this expression is $\sigma_{im} = \mathrm{E}\, R_i^e/k^*$. For the v_m portfolio, this gives $\sigma_m^2 = \mathrm{E}\, R_m^e/k^*$. Combine the last two expressions as $(\sigma_{im}/\sigma_m^2)\, \mathrm{E}\, R_m^e = \mathrm{E}\, R_i^e$. $\square$

Remark 6.3 *(Expected return of the tangency portfolio*) It follows directly that* $\mathrm{E}\, R_T^e = k\sigma_T^2$. *(Multiply both sides of (6.1) by* w_T*.) This says that the risk premium on the tangency portfolio increases if the risk aversion or variance does.*

6.1.2 Betas of Portfolios

Recall that the beta of any portfolio (not just the optimal one) is the weighted average (portfolio) of the betas of its components. That is, the portfolio with return

$$R_p = w'R^e + R_f \text{ has the beta} \tag{6.8}$$

$$\beta_p = w'\beta. \tag{6.9}$$

(This follows directly from $\beta_p = \mathrm{Cov}(\Sigma_{i=1}^n w_i R_i, R_T)/\sigma_T^2 = \Sigma_{i=1}^n w_i \beta_i$.)

Example 6.4 *Let* $(\beta_1, \beta_2) = (1.2, 0.8)$. *The portfolio return* $R_p = 0.6R_1 + 0.4R_2$ *has the beta* $\beta_p = 0.6 \times 1.2 + 0.4 \times 0.8 = 1.04$.

In particular, consider the portfolios on the capital market line (CML): $R_{opt} = vR_T^e + R_f$, where v is the weight on the tangency portfolio. Using the result in (6.9) and noticing that the tangency portfolio has $\beta_T = 1$ gives that $\beta = v$ for any portfolio on the CML.This implies that it is straightforward to create a portfolio with any desired β: just invest β in the tangency portfolio and $1 - \beta$ in the risk-free.

Example 6.5 *(Creating a portfolio with* $\beta_p = 0.44$*) We can create a portfolio with* $\beta = 0.44$ *by investing 0.44 in the tangency portfolio and 0.56 in the risk-free.*

6.1.3 Beta Representation and the Capital Market Line

The beta representation (6.5) means that two assets with the same betas should have the same expected returns—even if they have very different volatilities.

To be precise, consider the regression (6.7) which has the usual property that the residual is uncorrelated with the regressor. We can therefore write the variance of R_i as

$$\sigma_i^2 = \beta_i^2\sigma_T^2 + \sigma_\varepsilon^2. \tag{6.10}$$

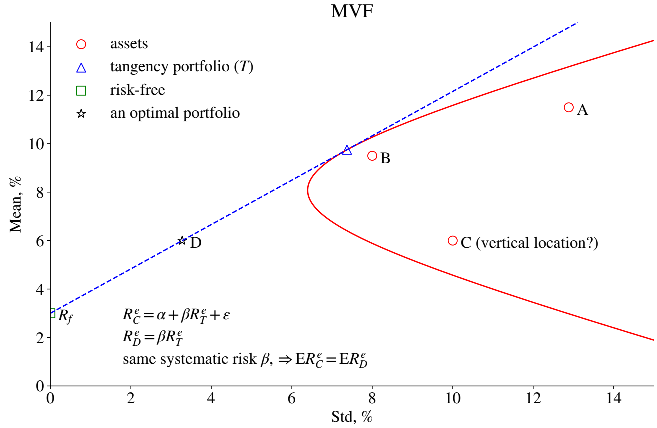


Figure 6.2: Mean-variance frontier and expected returns. The properties of the investable assets (A, B, and C) are shown in Table 6.1.

This says that the variance of R_i has two components: *systematic risk* (the comovement of R_i with R_T, $\beta_i^2\sigma_T^2$) and *idiosyncratic noise* (the variance of ε_i, σ_ε^2). In particular, *MV efficient portfolios have only systematic risk* ($\sigma_\varepsilon^2 = 0$) since they are formed from the tangency portfolio and risk-free ($R_{opt} = vR_T + (1-v)R_f$). All other portfolios with the same β are to the right in the MV figure: see Figure 6.2 for an illustration.

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 the beta representation, C must then have the same average return as D.*

6.1.4 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, which we assume is 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 driven by demand, at least in the short run.

Suppose all investors have the same beliefs about the asset returns (same expected returns and covariance matrix). They will then all mix the same tangency portfolio with the risk-free—but possibly in different proportions due to different risk aversions.

In equilibrium, net supply of the risk-free assets is zero (lending = borrowing), so *the average investor must hold the tangency portfolio* and no risk-free assets. Therefore, *the tangency portfolio must be the market portfolio*, so we can replace R_T with R_m in all expressions above. Therefore, CAPM says

$$\mathrm{E}\, R_i = R_f + \beta_i \mu_m^e \text{ where } \beta_i = \sigma_{im}/\sigma_m^2. \tag{6.11}$$

As discussed before, this expression is just a characterisation of the equilibrium (the first order conditions), and CAPM is silent on how that equilibrium is reached. One possible *story* is that β_i is driven by the firm characteristics (industry, size, leverage, etc.) and that equilibrium is reached as follows: high β assets are in low demand since they are too procyclical (pay off at the wrong time) which means that (in equilibrium) the share price will be low. For a given dividend stream, this means a higher dividend/price ratio, which contributes to a high average return.

Clearly, *CAPM relies on very strong assumptions*, in particular, the assumption about all investors having the same beliefs. Also, it rules out that investors face other types of financial risks (not just asset market risks). These issues will be discussed in later chapters.

6.1.5 Summarizing MV and CAPM: CML and SML

According to MV analysis, and assuming that the market portfolio equals the tangency portfolio, average return of all optimal (effective) portfolios (denoted opt) obey

$$\mathrm{E}\, R_{opt} = R_f + \sigma_{opt} SR_m. \tag{6.12}$$

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

In contrast, according to CAPM, the average return on all portfolios (optimal or not), obey the beta representation (6.6)

$$\mathrm{E}\, R_i = R_f + \beta_i \mu_m^e. \tag{6.13}$$

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

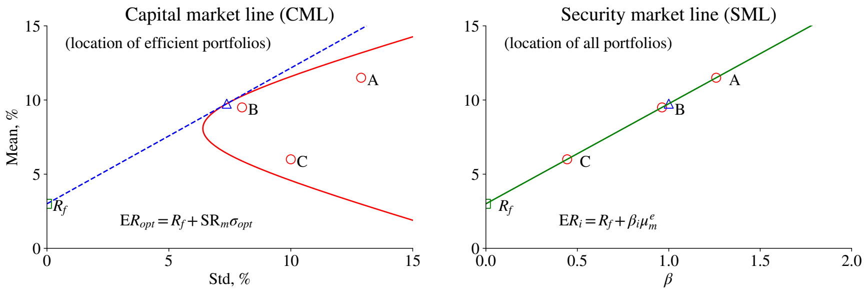


Figure 6.3: CML and SML. The properties of the investable assets (A, B, and C) are shown in Table 6.1.

6.2 More Properties of CAPM

6.2.1 Heterogeneous Beliefs*

This section discusses different risk aversions and heterogeneous beliefs about the mean returns. Both these allow for fairly straightforward aggregation. In contrast, handling different beliefs about the variance-covariance matrix are much harder.

We further assume that all investors have the same capital to invest (or more generally, if the capital of a investor is uncorrelated to the beliefs and risk aversion). Let superscript z to indicate the investor, so $\mathrm{E}^z\, R^e$ is the vector of his/her expected returns of the n assets and k^z is the risk aversion. Then, define the consensus (average) expectations across the Z investors as

$$\mathrm{E}^*\, R^e = \Sigma_{z=1}^{Z}\, \mathrm{E}^z\, R^e k^*/(k^z Z), \text{ where } k^* = 1/[\Sigma_{z=1}^{T} 1/(k^z Z)]. \tag{6.14}$$

Notice that k^* can be interpreted as an aggregate risk aversion. In the average expectations, investors with high risk aversion (and thus less inclined to invest in risky assets) have less weight. It follows that the market portfolio is the optimal portfolio ($v_m = \Sigma^{-1}\, \mathrm{E}^*\, R^e/k^*$) for an investor with the expectations in (6.14).

This gives a beta equation for consensus beliefs

$$\mathrm{E}^*\, R_i^e = \beta_i\, \mathrm{E}^*\, R_m^e, \tag{6.15}$$

where $\mathrm{E}^*\, R_i^e$ is an average ("consensus") expectation of the excess return on asset i, and m signifies the market portfolio. (See below for a proof.) Overall, this analysis suggests that CAPM might hold, at least as an approximation, for some types of heterogeneous beliefs.

Proof of (6.14)–(6.15). For investor z the first order conditions (6.1) are $\mathrm{E}^z R^e/k^z = \Sigma v^z$. Beliefs about the variance-covariance matrix Σ are assumed to be shared by all investors. Average the foc across the Z investors, $\Sigma_{z=1}^{Z} \Sigma v^z/Z$, to get $\Sigma v_m = \Sigma_{z=1}^{Z} \mathrm{E}^z R^e/(k^z Z)$, where v_m is the market portfolio. Define an aggregate risk aversion as in (6.14), and a consensus expectation as in the same equation. Then, the previous equation can be written $\Sigma v_m = \mathrm{E}^* R^e/k^*$, which can be solved for v_m. Notice that $\mathrm{E}^* R^e$ is a consensus (or average) belief. Row i of this expression is $\sigma_{im} = \mathrm{E}^* R_i^e/k^*$, since row i of Σv_m is the covariance of asset i and the v_m portfolio, σ_{im}. For the v_m portfolio, this gives $\sigma_m^2 = \mathrm{E}^* R_m^e/k^*$. Combine the last two expressions as $(\sigma_{im}/\sigma_m^2) \mathrm{E}^* R_m^e = \mathrm{E}^* R_i^e$. □

6.2.2 CAPM and Stochastic Discount Factors*

For future reference, we here notice that the CAPM expression (6.11)) can also be written in terms of a "stochastic discount factor" (SDF) model. This model implies

$$\mathrm{E}\, R_i^e M = 0, \text{ where } M = a - bR_m^e, \text{ with } b > 0. \tag{6.16}$$

Many asset pricing models can be written on a similar form, as will be discussed in later chapters.

Proof of (6.16) giving (6.11). Recall that $\mathrm{Cov}(x, y) = \mathrm{E}\, xy - \mathrm{E}\, x \times \mathrm{E}\, y$, so $\mathrm{E}\, xy = 0$ can be rearranged as $\mathrm{E}\, y = -\mathrm{Cov}(x, y)/\mathrm{E}\, x$. Applying to (6.16) gives

$$\mathrm{E}\, R_i^e = b\sigma_{im}/(a - b\,\mathrm{E}\, R_m^e)$$

We can, of course, apply this expression to the market excess return (instead of asset i) to get

$$\mathrm{E}\, R_m^e = b\sigma_m^2/(a - b\,\mathrm{E}\, R_m^e).$$

Solve for $b/(a - b\,\mathrm{E}\, R_m^e)$ and use that in the first equation to get the CAMP expression (6.11). □

6.2.3 Back to Prices (The Gordon Model)*

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

$$1 + R_{t+1} = (D_{t+1} + P_{t+1})/P_t, \tag{6.17}$$

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%) after period $t+1$, then it is straightforward to show that the asset price is

$$P_t = \mathrm{E}_t\, D_{t+1}/(R-g). \tag{6.18}$$

(See a later appendix on discounted cash flow.) 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 risk-free rate. One way of interpreting this is as follows. Suppose the asset (suddenly) gets a lower beta, which means that it has less systematic risk than before. It is therefore more useful in portfolio formation and thus 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). The valuation in (6.18) and CAPM are then consistent.

6.3 Testing CAPM

6.3.1 Testing a Single Asset

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 (6.11). This could be tested by the regression (6.7), but where we use the market return to proxy for the tangency portfolio return.

In particular, take average (over time) of the regression to get

$$\bar{R}^e_i = \hat{\alpha}_i + \hat{\beta}_i \bar{R}^e_m, \tag{6.19}$$

where $\bar{R}^e_i$ is the average excess return on asset i in the sample ($\bar{R}^e_i = \Sigma_{t=1}^T R^e_{it}/T$). Notice that $\bar{\varepsilon}_i = 0$ by construction.

The OLS estimate of β_i is the sample analogue to the true β_i. It is then clear that *CAPM implies* that the intercept (α_i) of the regression should be zero, which is what empirical tests of CAPM focus on.

However, this interpretation relies on a big *assumption*: that market expectations are (on average) well represented by the sample data. A rejection of CAPM in the test could therefore be driven either by some false assumptions behind CAPM or that the sample is systematically different from the market beliefs during the same period.

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

$$\frac{\hat{\alpha}_i}{\text{Std}(\hat{\alpha}_i)} \xrightarrow{d} N(0,1) \text{ under } H_0 : \alpha_i = 0. \tag{6.20}$$

In this expression, $\hat{\alpha}_i$ is the estimate of the intercept in (6.7) and $\text{Std}(\hat{\alpha}_i)$ its standard deviation (error), for instance, from the usual Gauss-Markov results for OLS. See also Remark 6.8 for a discussion of heterosekdasticity and autocorrelation. We reject the null hypothesis ($\alpha_i = 0$) when the the t-statistic is very negative or very positive (for instance, lower than -1.64 or higher than 1.64 for the 10% significance level, and $-1.96/1.96$ for the 5% level).

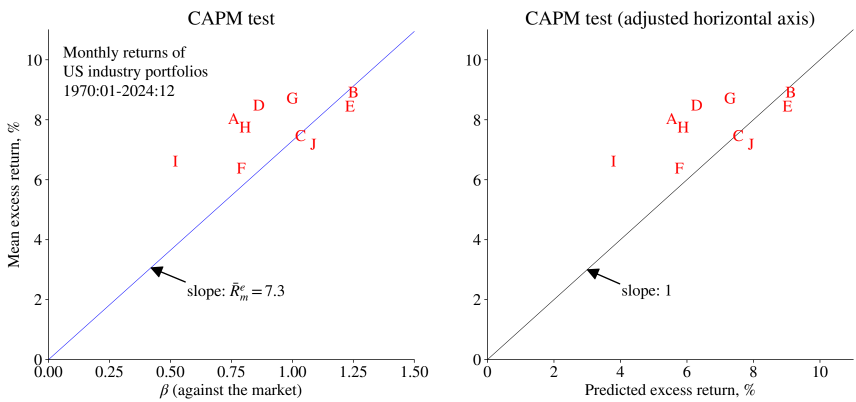
The test assets are often portfolios of firms with similar characteristics, for instance, small size or having their main operations in the retail industry. There are several reasons for testing the model on such portfolios: individual stocks are extremely volatile and firms can change substantially 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 extended.

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).

Empirical Example 6.7 *Figure 6.4 shows some results for US industry portfolios, while Table 6.2 and Figures 6.5–6.6 for US size/book-to-market portfolios. In these figures, the results are plotted in one of two different ways:*

$$\begin{array}{lll} & \underline{\textit{horizontal axis}} & \underline{\textit{vertical axis}} \\ 1: & \beta_i & \bar{R}_i^e \\ 2: & \beta_i \bar{R}_m^e & \bar{R}_i^e, \end{array} \tag{6.21}$$

where $\bar{R}_i^e$ indicates the (time) average excess return of asset i. In the first approach, CAPM says that all data points (different assets, i) should cluster around a straight line with a slope equal to the average market excess return, $\bar{R}_m^e$. 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



	α (ann.)	t-stat	σ (ann.)
A (NoDur)	2.44	2.06	8.73
B (Durbl)	-0.25	-0.11	17.26
C (Manuf)	-0.12	-0.14	6.43
D (Enrgy)	2.13	0.93	16.84
E (HiTec)	-0.63	-0.41	11.17
F (Telcm)	0.57	0.38	10.99
G (Shops)	1.38	1.13	8.97
H (Hlth)	1.81	1.22	10.94
I (Utils)	2.77	1.74	11.75
J (Other)	-0.80	-0.85	6.97

CAPM: $R^e_i = \alpha_i + \beta_i R^e_m + e_i$
Predicted excess return: $\beta_i \bar{R}^e_m$

10% crit. value (Bonferroni): 2.58

Test if all $\alpha_i = 0$:

Wald stat	10.55
5% crit val	18.31
p-value	0.39

Figure 6.4: CAPM regressions on US industry indices

to CAPM).

6.3.2 Testing 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, because that is the implication of CAPM. Ideally we then want to take into account the correlation of the different alpha estimates.

While it is straightforward to construct such a test, it requires setting up a system of regression equations and test across regressions. See Remark 6.8 for details. Alternatively, we can apply a *Bonferroni adjustment* of the individual t-stats: reject CAPM at the 10% significance level only if the largest t-stat (in absolute terms) exceeds the critical value at the $0.10/n$ significance level. For instance, with $n = 25$, the critical value from a standard normal distribution would be 2.88 instead of 1.64. The motivation for this is that repeated single-asset testing (using a traditional critical value) will, by pure randomness, reject 10% of the cases—even if the null hypothesis is true. The Bonferroni adjustment takes this into account to correct the "family-wise" false rejection rate.

Remark 6.8 *(*Variance-covariance matrix if OLS estimate of* α*) The "system estimation" is actually* n *separate OLS regressions. Assuming residuals have no autocorrelation or heteroskedasticity (the standard Gauss-Markov assumption), it is straightforward to show that the variance-covariance matrix of the vector of the estimated* $\hat{\alpha}$*-values* $(\hat{\alpha}_1, \hat{\alpha}_2, ..., \hat{\alpha}_n)$ *is* $V = \Omega(1 + SR_m^2)/T$*, where* Ω *is the variance-covariance matrix of the residuals,* SR_m *is the Sharpe ratio of the market and* T *the (time) length of the sample. This holds also for a single asset as in (6.20). For monthly or longer return periods, autocorrelation is rarely a problem and a heteroskedasticity-robust* V *is typically similar (for* α*, not for* β*). The joint hypothesis that all alphas are zero can be tested with an* F*- or* χ^2*-test. The latter has the test statistic* $\hat{\alpha}' V^{-1} \hat{\alpha}$*, which is distributed as a* χ_n^2 *variable under the null hypothesis.*

	1	2	3	4	5
1	−3.45	−0.09	0.44	2.11	2.52
2	−2.33	0.53	1.43	2.30	1.84
3	−2.19	1.28	1.20	2.22	2.22
4	−0.86	0.39	1.21	2.06	1.54
5	0.24	1.23	1.17	0.11	0.97

Table 6.2: t-stats for α in CAPM, 25 FF portfolios 1970:01-2024:12. NW uses 1 lag. The Bonferroni adjusted 10% and 5% critical values are 2.88 and 3.09.

A quite different approach to study a cross-section of assets is to first perform a CAPM regression for each asset, and use the estimated betas as regressors in the following cross-sectional regression

$$\bar{R}_i^e = \gamma + \lambda \hat{\beta}_i + u_i, \tag{6.22}$$

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 estimate), which typically tend to bias the slope coefficient towards zero ("errors in variables"). 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 (6.22) to have a high R^2—since it should be unity in a very large sample (if CAPM holds).

6.3.3 Representative Results of the CAPM Test

One of the more interesting studies is Fama and French (1993) (see also Fama and French (1996)). They construct 25 stock portfolios according to two characteristics of the firm: the size (by market capitalization) and the book-value-to-market-value ratio (BE/ME).

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. However, it is found that there is almost no relation between $\bar{R}_i^e$ and β_i (there is a cloud in the $\beta_i \times \bar{R}_i^e$ 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 $\bar{R}_i^e$ and β_i—as predicted by CAPM. Second, *within a size quintile* there is a negative relation (across BE/ME quantiles) between $\hat{R}_i^e$ and β_i—in stark contrast to CAPM. See Figures 6.5–6.6 for results from more recent data.

6.3 Appendix – Discounted Cash Flow*

6.3.1 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 6.9 *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)

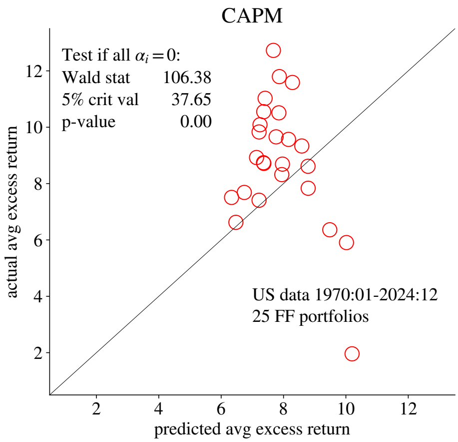


Figure 6.5: CAPM, FF portfolios

price equals the present value of future cash flow

$$P_t = \sum_{s=1}^{\infty} \frac{\mathrm{E}_t \, D_{t+s}}{(1+R)^s}, \tag{6.23}$$

where D_{t+s} are the future cash flows to the investor. In this expression subscripts refer to time periods.

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 here assumed to be constant). In general, the discount rate depends on both the risk-free rate and the risk of the asset. 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 6.10 *(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 $ROE = R$.

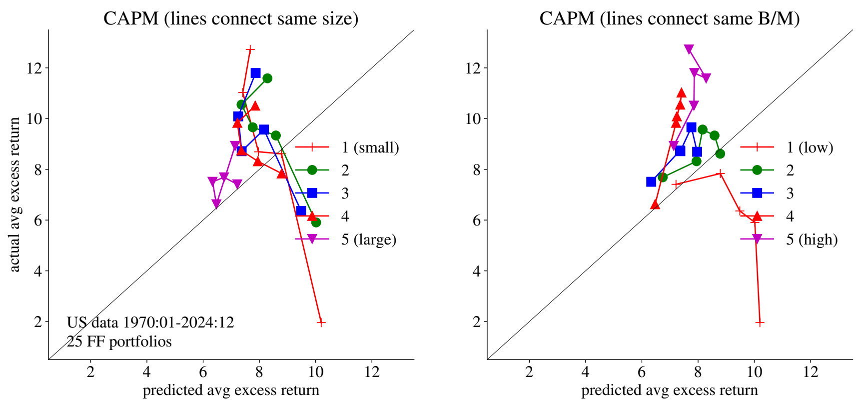


Figure 6.6: CAPM, FF portfolios

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 (Taxes, changes in risk and beliefs about future cash flows can complicate this story.) See Fabozzi, Neave, and Zhou (2012) for a discussion.

Remark 6.11 *(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.*

Remark 6.12 *(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*

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

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 6.13 *(From income to cash flow*) To calculate the cash flow start with the earnings before interests and taxes (EBIT) from the income statement, subtract taxes on EBIT (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 6.14 *(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.

6.3.2 "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

$$P_t = \frac{\mathrm{E}_t\, D_{t+1} + \mathrm{E}_t\, P_{t+1}}{1+R}. \tag{6.24}$$

This is the same as the fundamental valuation (6.23) *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 (6.23) to write

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

Take expectations as of period t and use in (6.24)

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

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 6.15 *(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.*

6.3.3 Fundamental Valuation and Returns

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

$$R_{t+1} = \frac{D_{t+1} + P_{t+1}}{P_t} - 1. \tag{6.25}$$

If the discount rate in (6.23) is constant over time, then it equals the expected return

$$\mathrm{E}_t\, R_{t+1} = R. \tag{6.26}$$

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

$$R_{t+1} = R \text{ (if no news)}. \tag{6.27}$$

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 (6.27)–(6.26). Use (6.23) to write

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

Use in the realized return (6.25) and take expectations as of t to get (using $\mathrm{E}_t\, \mathrm{E}_{t+1}\, D_{t+s} = \mathrm{E}_t\, D_{t+s}$)

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

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 (6.25).) □

6.3.4 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 (6.23) becomes

$$P_t = (\mathrm{E}_t\, D_{t+1})\textstyle\sum_{s=1}^{\infty}\frac{(1+g)^{s-1}}{(1+R)^s} = \frac{\mathrm{E}_t\, D_{t+1}}{R-g}. \tag{6.28}$$

This is the "Gordon model." The asset price (6.28) 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")

$$R = \frac{\mathrm{E}_t\, D_{t+1}}{P_t} + g. \tag{6.29}$$

Example 6.16 *(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 (6.28) Write the first equality of (6.28) 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

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

□

6.3.5 Valuation Multiples

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

$$\text{“}p/e\text{”} = \frac{P}{e}, \tag{6.30}$$

where e is short for earnings per share.

If dividends are proportional to earnings, $D_t = k \times e_t$ in each period and earnings grow at the rate g, $e_{t+1} = (1+g)e_t$, then

$$(p/e)_t = \frac{P_t}{e_t} = \frac{\overbrace{ke_{t+1}}^{D_{t+1}}/(R-g)}{e_t} = k\frac{1+g}{R-g}.$$

Example 6.17 *$R = 0.1$, $g = 2\%$ and $k = 1$ (a "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

$$P_{i,t} = \left(\frac{P_t}{e_t}\right)_{peers} \times e_{i,t}, \text{ so } \frac{P_{i,t}}{e_{i,t}} = \left(\frac{P_t}{e_t}\right)_{peers}. \tag{6.31}$$

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 6.18 *(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 (6.31), 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, or somehow counterbalance each other.

Chapter 7

Downside Risk Measures

The mean-variance framework is often criticized for failing to distinguish between the downside of the return distribution (considered to represent risk) and upside (considered to represent potential). This chapter introduces several commonly used downside risk measures and discusses to what extent this leads to a different the ranking of assets. (Later chapters will embedded these new measures into portfolio choice models.)

7.1 Value at Risk

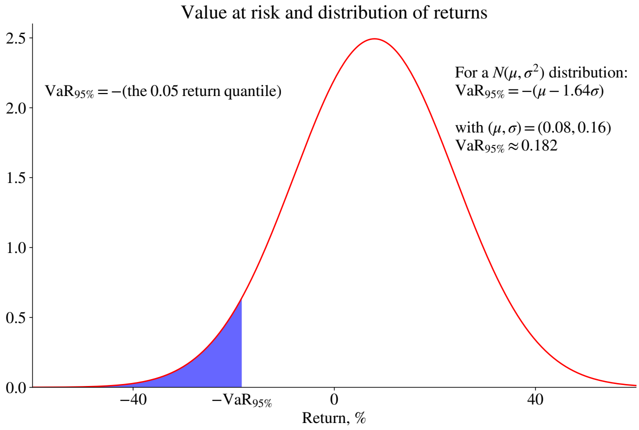


Figure 7.1: Value at risk

The Value at Risk (VaR) measures the downside risk by focusing on a quantile of the return (or loss) distribution,

Remark 7.1 *(Quantile of a distribution) The 0.05 quantile is the value x such that there is only a 5% probability of a lower number,* $\Pr(R \leq x) = 0.05$.

The 95% Value at Risk ($\text{VaR}_{95\%}$) is a number such that there is only a 5% chance that the loss as a fraction of the investment (which is the negative of the return, $-R$) is larger

$$\Pr(-R \geq \text{VaR}_{95\%}) = 5\%. \tag{7.1}$$

Here, 95% is the confidence level of the VaR. For instance, with $\text{VaR}_{95\%} = 18\%$, then we are 95% sure that we will not lose more than 18% of our investment.

To work with the return distribution, not the loss distribution, we notice that (7.1) is the same as

$$\Pr(R \leq -\text{VaR}_{95\%}) = 5\%, \tag{7.2}$$

so $-\text{VaR}_{95\%}$ is the 0.05 *quantile* of the return distribution. More generally, the VaR for confidence level α (instead of 0.95) is

$$\text{VaR}_{\alpha} = -(\text{the } 1-\alpha \text{ quantile of the } R \text{ distribution}). \tag{7.3}$$

If the return is *normally distributed*, $R \sim N(\mu, \sigma^2)$, then

$$\text{VaR}_{\alpha} = -(\mu + c\sigma), \tag{7.4}$$

where c is the $1-\alpha$ quantile of a $N(0,1)$ distribution. For instance, c is approximately $(-1.64, -1.96, -2.33)$ for the $(0.05, 0.025, 0.01)$ levels, respectively. See Figures 7.1–7.2. Since $c < 0$, the VaR is here strictly increasing the the standard deviation, which will later be important when we consider portfolio choice based on a VaR.

Example 7.2 *(VaR and regulation of bank capital) Bank regulations have used 3 times the 99% VaR for 10-day returns as the required bank capital.*

Note that the return distribution depends on the *investment horizon*; therefore, the VaR is typically calculated for a particular return period (for instance, one day). Multi-period VaRs are calculated by either explicitly constructing the distribution of multi-period returns, or by making simplifying assumptions about the relation between returns in different periods (for instance, that they are uncorrelated). 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.

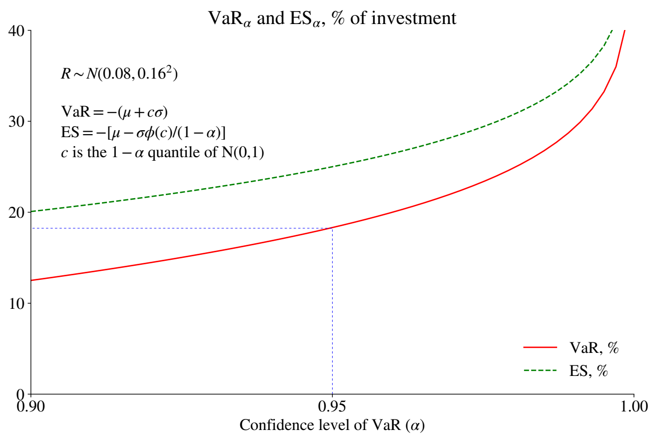


Figure 7.2: Value at risk and expected shortfall, different confidence levels

Example 7.3 *(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 credit derivatives. However, (a) that would have caused high execution costs and (b) the portfolio had recently performed well. At this time a new VaR method was developed and quickly implemented. The division went on to triple the positions (and lose $719 million in 2012Q1). Interestingly, the two VaR models (old and new) show divergent paths for the value at risk, with the new suggesting much lower risk.*

To use VaR as a risk control, or more generally, in portfolio choice, we first need to formulate beliefs (estimates) of the future return distribution. The next section is aimed at assessing how well some approaches do in capturing the future (ex post) return movements.

7.1.1 Backtesting a VaR model

Backtesting a VaR model amounts to comparing how well the VaR model can describe the 5% quantile (say) of the ex post data. This is done by determining whether the returns fall below the VaR$_{95\%}$ approximately 5% of the times in the sample. This could be a long sample period or a set of subsamples. In particular, a model with extended periods of under- and then over-performance which average (in the full sample) to roughly 5% is unlikely to be a useful model.

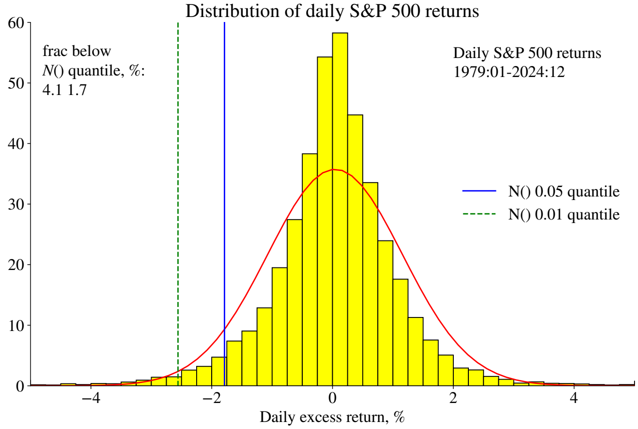


Figure 7.3: Return distribution and VaR for S&P 500

Empirical Example 7.4 *Figure 7.3 shows the distribution and static VaRs (using constant parameters) for the daily S&P 500 returns for a long sample. The results indicate that the $N()$-based model has a reasonable coverage for the 95%, but perhaps not for the 99% confidence level.*

Empirical Example 7.5 *Figure 7.4 shows backtesting on many subsamples. The results indicate that a static VaR model for S&P 500 has long cycles of under- and then over-performance.*

7.1.2 A Simple Dynamic VaR

It is well known that financial *volatility changes over time*, which needs to be embedded in a reliable VaR model. One particularly simple approach is to estimate means and variances, using the recursive formulas (the RiskMetrics approach, see JP Morgan (1996))

$$\mu_t = \lambda \mu_{t-1} + (1-\lambda) R_{t-1} \tag{7.5}$$

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1-\lambda)(R_{t-1} - \mu_{t-1})^2, \tag{7.6}$$

where $0 < \lambda < 1$ and often high (around $0.90 - 0.95$ for daily data). The estimate of the mean is an update of yesterday's estimate, using yesterday's return for the update. This is the same as a weighted average of past returns (actually, an exponentially

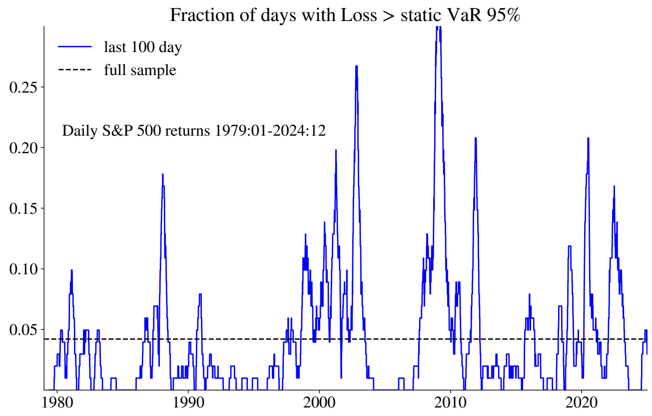


Figure 7.4: Backtesting a static VaR model on a moving data window

weighted moving average, EWMA), with recent data having higher weights than old data. The variance is a similar, with updating using the square of yesterday's surprise.

Empirical Example 7.6 *Figure 7.5 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. Figures 7.5–7.6 show results from backtesting a VaR model which assumes that one-day returns are normally distributed, but where the mean and volatility are time varying. 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.*

7.1.3 Value at Risk of a Portfolio

The general way of calculating the VaR of a portfolio is the same as for an individual asset: first calculate (or estimate) the parameters of the distribution, then find the quantile.

However, in some special cases, it is possible to directly translate the VaR values of the individual assets into a portfolio VaR.

Remark 7.7 *Suppose the assets in the portfolio are jointly normally distributed with zero means, so the VaR of asset i is $VaR_i = 1.64\sigma_i$. (The index on VaR here*

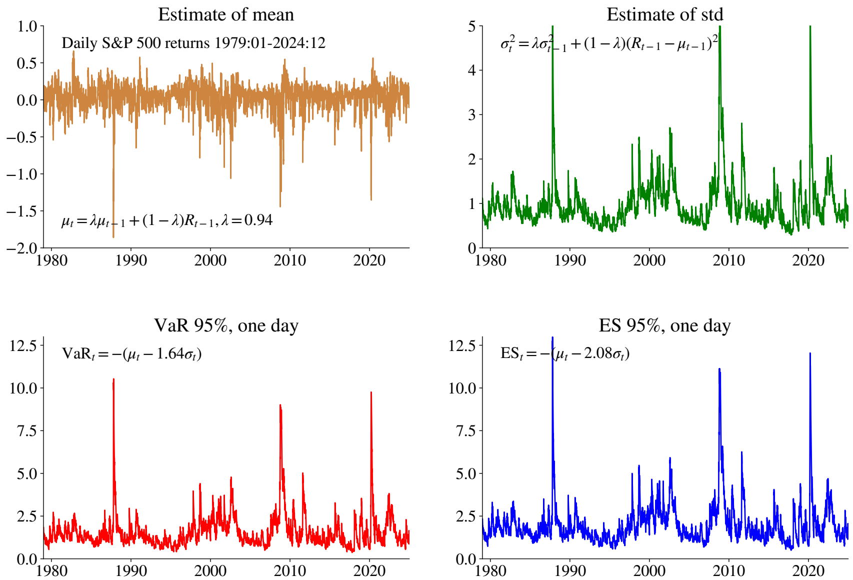


Figure 7.5: A dynamic VaR model

indicates the asset, not a confidence level.) 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 see this, recall that* $VaR_p = 1.64\sigma_p$ *and that we can calculate* σ_p *from the* σ_i *values and correlations.)*

7.1.4 Index Models for Calculating the Value at Risk

Consider a multi-index model

$$R_t = a + b' I_t + e_t, \tag{7.7}$$

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

$$\begin{aligned} \mu &= a + b' \mathrm{E}\, I_t \\ \sigma &= \sqrt{b' \operatorname{Cov}(I_t) b + \operatorname{Var}(e_t)}, \end{aligned} \tag{7.8}$$

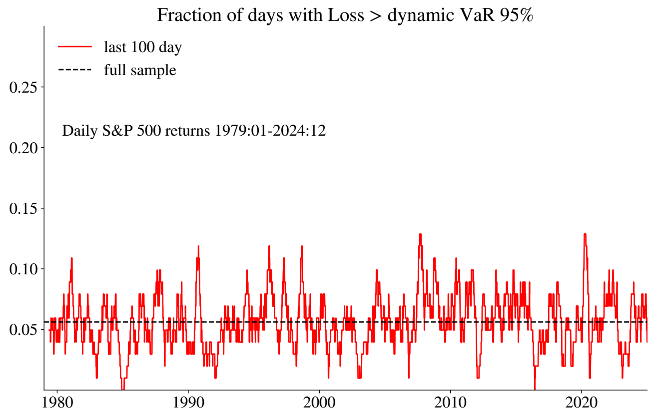


Figure 7.6: Backtesting a dynamic VaR model on a moving data window

which can be used in (7.4). If the return is of a well diversified portfolio and the indices include the key market indices, then the idiosyncratic risk $\mathrm{Var}(e)$ is close to zero. The RiskMetrics approach is to make this assumption.

A *stand-alone VaR* assesses the contribution of different factors (indices) on the overall VaR. For instance, the indices in (7.7) could include: 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 (7.7) 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. This approach, when combined with the normal assumption in (7.4), is called the *delta-normal method*.

7.2 Expected Shortfall

While the value at risk is a useful risk measure, it has the strange property that it only considers whether an 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 of a portfolio is not necessarily (weakly) lower than the portfolio of the VaRs even if the assets all have the same volatility, which contradicts the notion of diversification benefits. (To get this unfortunate property, the return distributions must be heavily skewed.) See McNeil, Frey, and Embrechts (2005) and Alexander (2008) for more detailed discussions.

The expected shortfall (ES, 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,

$$\mathrm{ES}_\alpha = -\mathrm{E}(R|R \leq -\mathrm{VaR}_\alpha). \tag{7.9}$$

See Figure 7.7 for an illustration.

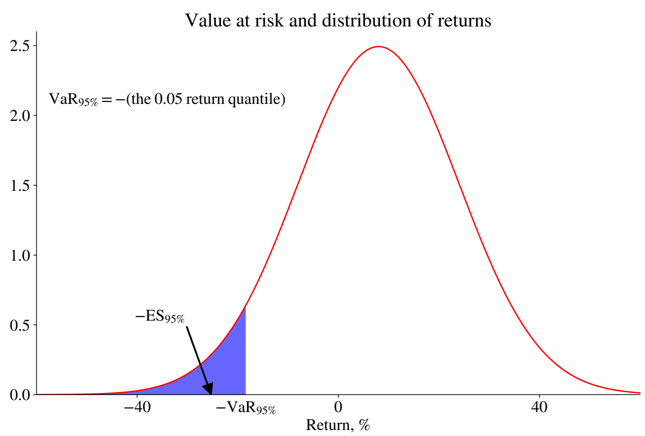


Figure 7.7: Value at risk and expected shortfall

Empirical Example 7.8 *See Figure 7.5 for an empirical estimate of ES, based on the dynamic estimates of the mean and variance in (7.5)–(7.6).*

Empirical Example 7.9 *See Table 7.1 for an empirical comparison of the VaR,* *ES and some alternative downside risk measures (discussed below) for two stock indices.*

	Small growth	Large value
Std	8.1	5.8
VaR (95%)	12.9	8.8
ES (95%)	17.9	13.1
SemiStd	5.7	3.9
Drawdown	78.4	63.2

Table 7.1: Risk measures of monthly returns of two stock indices (%), US data 1970:01-2024:12.

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

$$\text{ES}_\alpha = -[\mu - \phi(c)\sigma/(1-\alpha)], \tag{7.10}$$

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, with $1-\alpha = 0.05$ and thus $c \approx -1.64$. In this case, the ES is strictly increasing in the standard deviation, which will later be important when we consider portfolio choice. See Figure 7.2 for an illustration.

Example 7.10 *(ES) If $\mu = 8\%$ and $\sigma = 16\%$, the 95% expected shortfall is $ES_{95\%} = -(0.08 - 2.08 \times 0.16) \approx 0.25$ (since $\phi(-1.64)/0.05 \approx 2.08$) and the 97.5% expected shortfall is $ES_{97.5\%} = -(0.08 - 2.34 \times 0.16) \approx 0.29$ (since $\phi(-1.96)/0.025 \approx 2.34$)*

Proof of (7.10). 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. $\square$

To estimate the average shortfall from a sample, calculate the average $-R_t$ for observations where $R_t \leq -\text{VaR}_\alpha$

$$\text{ES}_\alpha = -\textstyle\sum_{t=1}^T \delta_t R_t / (\Sigma_{t=1}^T \delta_t), \text{ where } \delta_t = \delta(R_t \leq -\text{VaR}_\alpha) \tag{7.11}$$

and where $\delta(q) = 1$ if q is true and zero otherwise. (In this expression δ_t is a dummy variable whose value depends on the $\delta()$ function.) This can be used in backtesting an ES model.

Empirical Example 7.11 *See Figure 7.8 for a back testing of the dynamic ES model previously shown in Figure 7.5.*

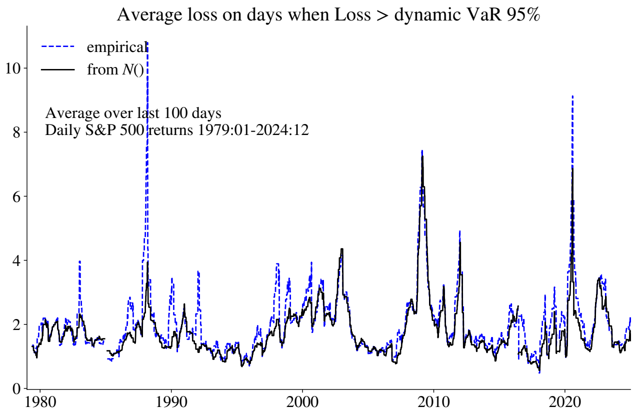


Figure 7.8: Backtesting a dynamic ES model on a moving data window

7.3 Target Semi-Variance

The target semi-variance (TSV, also called the lower partial 2nd moment) is defined as

$$\sigma_{-}^{2}(h) = \mathrm{E}[\min(R-h,0)^{2}], \tag{7.12}$$

where h is a "target level" chosen by the investor. Also, $\sqrt{\sigma_{-}^{2}(\mu)}$ is called the semi-standard deviation. In comparison with the variance, $\sigma^2 = \mathrm{E}(R - \mathrm{E}\,R)^2$, the target semi-variance 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 (see Bawa and Lindenberg (1977) and Nantell and Price (1979)).

For a normally distributed variable, the target semi-variance $\sigma_{-}^{2}(h)$ is increasing in the standard deviation, see Remark 7.13, which will later be important when we consider portfolio choice. See also Figure 7.9 for an illustration.

To estimate the target semi-variance from the empirical return distribution (for backtesting), use

$$\sigma_{-}^{2}(h) = \frac{1}{T}\sum\nolimits_{t=1}^{T}\delta_t(R_t-h)^2, \text{ where } \delta_t = \delta(R_t \le h) \tag{7.13}$$

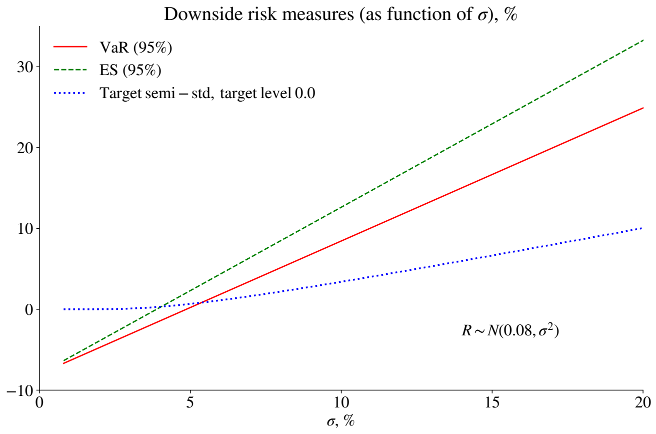


Figure 7.9: Downside risk measures as functions of the standard deviation for a N(μ,σ^2) variable

and where $\delta(q) = 1$ if q is true and zero otherwise.

Remark 7.12 *(*Alternative scaling of $\sigma_-^2(h)$) Some analysts define $\sigma_-^2(h)$ by just including those observations for which $R_t \leq h$. This means multiplying (7.13) by $T/\Sigma_{t=1}^T \delta(R_t \leq h)$, which is actually estimating* $\mathrm{E}[(R-h)^2 | R_t \leq h]$.

Remark 7.13 *(Target semi-variance calculation for normally distributed variable*) For an $N(\mu,\sigma^2)$ variable, target semi-variance around the target level h is*

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

where $\phi()$ and $\Phi()$ are the pdf and cdf of a $N(0,1)$ variable, respectively. Notice that $\sigma_-^2(\mu) = \sigma^2/2$. It is straightforward to show that $d\sigma_-^2(h)/d\sigma = 2\sigma\Phi(a)$, so the target semi-variance is a strictly increasing function of the standard deviation.

Remark 7.14 *(Sortino ratio) The Sortino ratio is an alternative to the Sharpe ratio as a measure of performance. It is* $(\mathrm{E}\,R - h)/\sqrt{\sigma_-^2(h)}$.

Empirical Example 7.15 *See Table 7.2 for an empirical rank correlation of the different risk measures for 25 FF portfolios. Most of the risk measures have strong rank correlations, meaning that they give very similar ranking of "riskiness" of these 25 assets. However, max drawdown is different, mostly likely since it is focused on the extreme left tail of the distribution.*

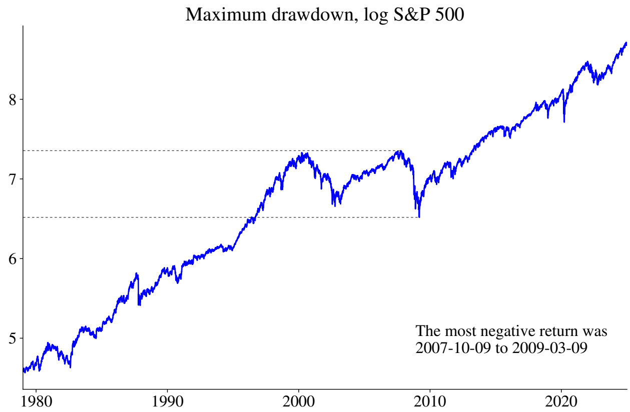


Figure 7.10: Maximum drawdown over the full sample

	Std	VaR (95%)	ES (95%)	SemiStd	Drawdown
Std	1.00	0.95	0.97	0.98	0.61
VaR (95%)	0.95	1.00	0.95	0.96	0.65
ES (95%)	0.97	0.95	1.00	0.97	0.64
SemiStd	0.98	0.96	0.97	1.00	0.60
Drawdown	0.61	0.65	0.64	0.60	1.00

Table 7.2: Correlation of rank of risk measures across the 25 FF portfolios (%), US data 1970:01-2024:12. The VaR and ES are based on the empirical return distribution. The max drawdown is calculated over a moving 5-year data window.

7.4 Maximum Drawdown

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 7.10. 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.

Empirical Example 7.16 *See Figure 7.11 for a comparison of the max drawdown of two return series. The results suggest that small growth stocks are considerably more risky than large value stocks.*

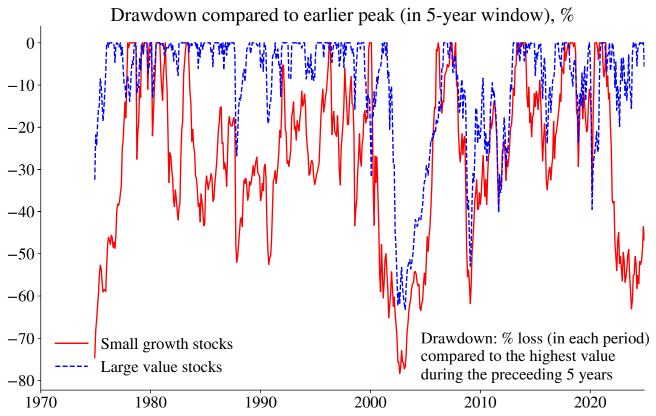


Figure 7.11: Drawdown

7.5 Empirical Return Distributions

Are returns normally distributed? Mostly not, but it depends on the asset type and on the data frequency. Options returns have very non-normal distributions (in particular, since the return is -100% on many expiration days). Stock returns are typically distinctly non-normal at short horizons, but may appear approximately normal over longer horizons. This may (or may not) carry over to the beliefs held by investors.

To assess the normality of realized 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.

Remark 7.17 *(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 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 (0.02 quantile) 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.*

Empirical Example 7.18 *See Figures 7.12–7.14 for empirical histograms and QQ-plots of S&P 500 returns. It is observed, among other findings, that empirical returns*

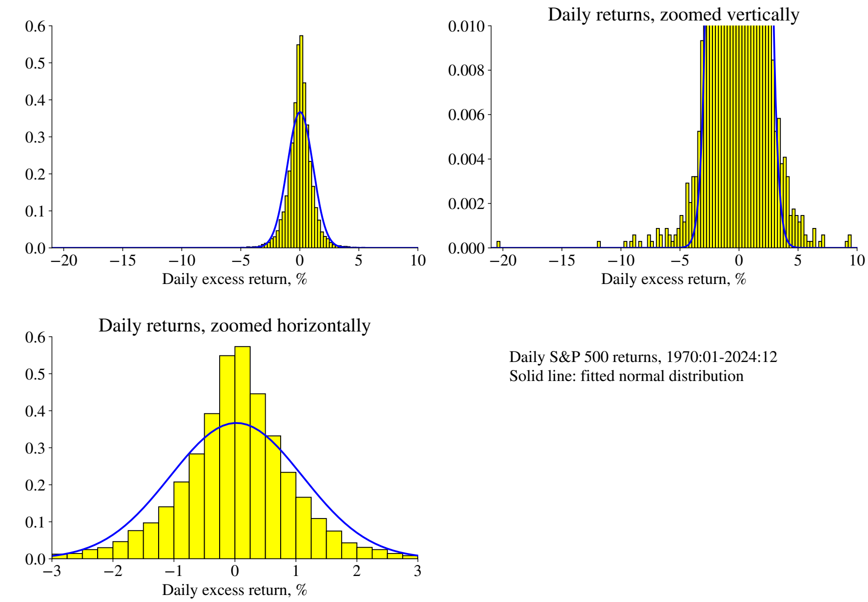


Figure 7.12: Distribution of daily S&P returns

distributions exhibit more extreme negative returns than a normal distribution would suggest, and that the return distribution looks closer to a normal distribution as the return horizon increases.

These methods can be applied to both data on returns and to residuals from a statistical model. For instance, the $(R_t - \mu_t)/\sigma_t$ where (μ_t, σ_t) are estimated by (7.5)–(7.6).

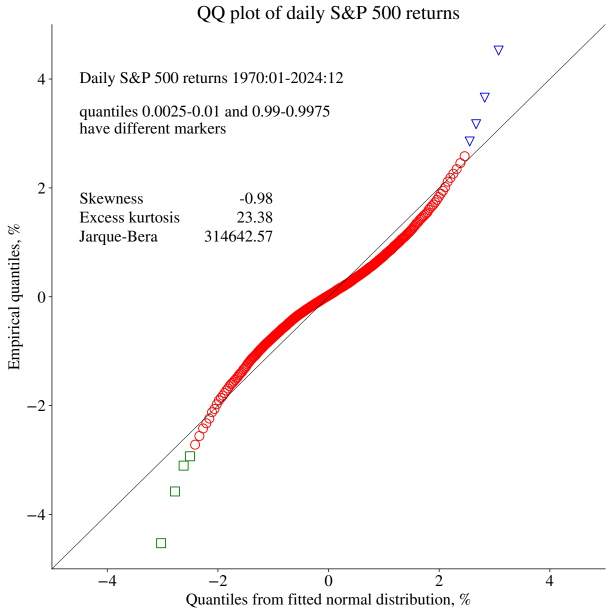


Figure 7.13: Quantiles of daily S&P returns

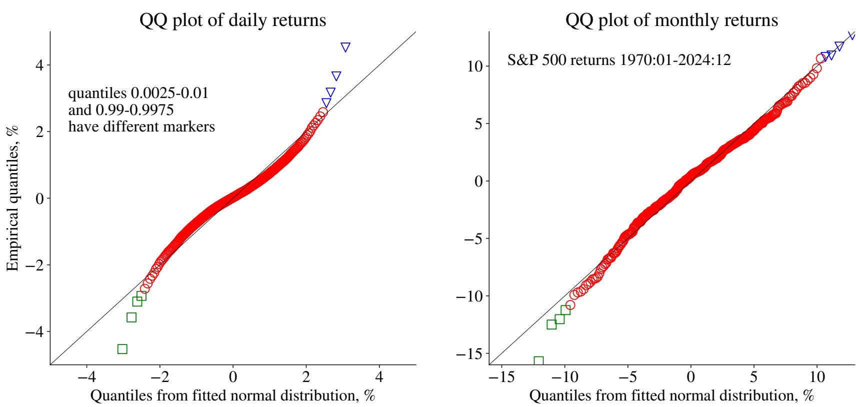


Figure 7.14: Distribution of S&P returns (different horizons)

Chapter 8

Utility-Based Portfolio Choice

Earlier chapters have assumed that investor preferences can be summarized by the mean and variance of portfolio returns. This chapter takes a step back to study whether that is consistent with established utility theory—and in some cases it is.

8.1 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 go beyond it—we must refer to economic utility theory.

8.1.1 Specification of Utility Functions

In finance, the key features of utility functions are as follows. *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 *real* (inflation adjusted) portfolio return. In one-period investment problems, this choice of x is irrelevant since consumption equals wealth, which is proportional to the portfolio return.

Second, uncertainty is incorporated by letting investors maximize expected utility, $\mathrm{E}\, U(x)$. The reasons is that returns (and therefore wealth and consumption) are uncertain. Hence, we need a way to rank portfolios at the time of investment, before the uncertainty is resolved. For instance, if there are S possible states with outcomes $x_1, x_2, ..., x_S$ and probabilities $\pi_1, \pi_2, ...\pi_S$, then expected utility is

$$\mathrm{E}\, U(x) = \textstyle\sum_{s=1}^{S} \pi_s U(x_s). \tag{8.1}$$

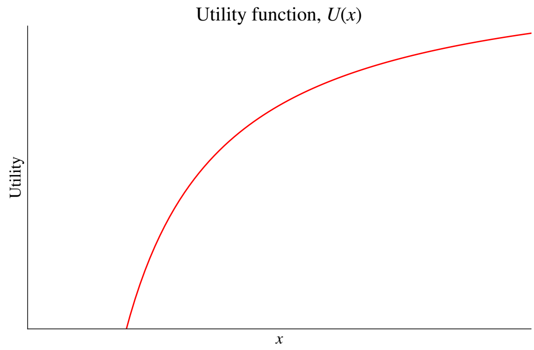


Figure 8.1: A utility function

The outcomes could represent portfolio returns, which depend on the state *and* the portfolio weights. That is, x_s is not a fixed list, but rather functions of the investor's choices. As usual, the expectation is based on the investor's beliefs.

Example 8.1 *(*$\mathrm{E}\,U(W)$*) Suppose there are two states of the world: W (wealth) will be either* 0.85 *or* 1.15 *with probabilities* $1/3$ *and* $2/3$*. If* $U(W) = \ln W$*, then* $\mathrm{E}\,U(W) = 1/3 \times \ln 0.85 + 2/3 \times \ln 1.15 \approx 0.039$*. If the investor had picked another portfolio the outcomes would instead be* 0.9 *and* 1.05*, with expected utility* $1/3 \times \ln 0.9 + 2/3 \times \ln 1.05 \approx -0.003$.

Third, the functional form of the utility function is such that more is better (the function is increasing) and the function is concave. The latter means that investors are risk averse. See Figure 8.1 for an illustration.

Remark 8.2 *(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 (a 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* (a) *X with prob π and Z with prob* $1-\pi$ *and* (b) *Y with prob π and Z 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).*

8.1.2 Basic Properties of Utility Functions: (1) More is Better

The idea that *more is better* (non-satiation) is 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 8.3 *(Logarithmic utility)* $U(W) = \ln W$ *so* $U'(W) = 1/W > 0$ *(assuming* $W > 0$*).*

8.1.3 Basic Properties of Utility Functions: (2) Risk is Bad

With expected utility, *risk aversion* (uncertainty is considered to be bad) is captured by the concavity of the utility function.

In contrast, a linear utility function implies risk-neutrality, which we rule out because investors appear to care about risk. (Some may seems not to do so, but they are often not gambling with their own money.)

As an example, consider Figure 8.2. It shows a case where the portfolio (or wealth, or consumption,...) of an investor will pay either x^- or x^+ with equal probabilities. The utility function embodies 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

$$U(\mathrm{E}\,x) > 0.5U(x^-) + 0.5U(x^+) = \mathrm{E}\,U(x). \tag{8.2}$$

Remark 8.4 *(*Risk aversion and "marginal utility") Rearranging (8.2) gives*

$$U(\mathrm{E}\,x) - U(x^-) > U(x^+) - U(\mathrm{E}\,x),$$

which says that moving from a low to a mid value of x (left hand side) counts for more than moving from a mid value to a high value (right hand side). Another way of phrasing 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 that gives the same utility as $\mathrm{E}\,U(x)$, that is, the value of P that solves the equation

$$U(P) = \mathrm{E}\,U(x). \tag{8.3}$$

This price (P), called the *certainty equivalent* of the portfolio, is less than the

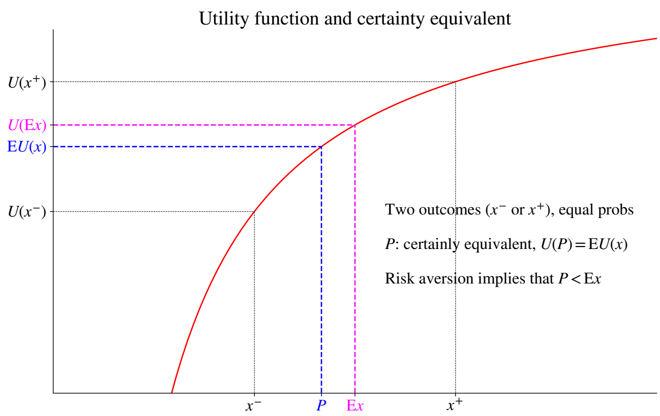


Figure 8.2: Certainty equivalent

expected payoff

$$P < \mathrm{E}\,x = 0.5x^{-} + 0.5x^{+}. \tag{8.4}$$

(The result follows from $U(P) < U(\mathrm{E}\,x)$ and that $U()$ is an increasing function.) Again, see Figure 8.2 for an illustration.

Example 8.5 *(Certainty equivalent) Suppose you have a CRRA utility function, $x^{1-\gamma}/(1-\gamma)$, 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*

$$\frac{P^{1-\gamma}}{1-\gamma} = 0.5\frac{0.85^{1-\gamma}}{1-\gamma} + 0.5\frac{1.15^{1-\gamma}}{1-\gamma}.$$

For instance, with $\gamma = 0, 2, 5, 10,$ and 25 we have $P \approx 1, 0.977, 0.947, 0.912,$ and 0.875. Notice that P goes from the average payoff (1) to the lowest outcome (0.85) as risk aversion increases.

This means that the *expected net return* on the risky portfolio that the investor demands is

$$\mathrm{E}\,R_x = \frac{\mathrm{E}\,x}{P} - 1 > 0, \tag{8.5}$$

which is greater than zero. This “required return” is higher if the investor is very risk averse (since P is lower). 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.

Example 8.6 *(Risk premium in a simple case) Using the* $k = 2$ *case in Example 8.5 we get the expected net return (8.5)* $1/0.977 - 1 \approx 2.4\%$, *since* $\mathrm{E}\,x = 1$. *Instead, with* $k = 25$ *we get* $1/0.875 - 1 \approx 14.3\%$.

8.2 Utility-Based Portfolio Choice and MV Frontiers

8.2.1 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 risk-free asset

$$\max_v \mathrm{E}\,U(R_p), \text{ with } R_p = vR^e + R_f, \tag{8.6}$$

where R^e denotes the excess return of the risky asset.

The first order condition with respect to the weight on risky assets is

$$\begin{aligned} \frac{d\,\mathrm{E}\,U(vR^e + R_f)}{dv} &= 0 \text{ or} \\ \mathrm{E}[U'(R_p)R^e] &= 0, \end{aligned} \tag{8.7}$$

where $U'(R_p)$ is the marginal utility evaluated at $R_p = vR^e + R_f$. 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 8.7 *(Stochastic discount factor models*) Equation (8.7) is on the same form as a stochastic discount factor (SDF) to asset pricing, where* $\mathrm{E}\,MR^e = 0$ *is a key condition.*

As an example, with a CRRA utility function the first order condition (8.7) can be written

$$\mathrm{E}\,\frac{R^e}{(vR^e + R_f)^\gamma} = 0, \tag{8.8}$$

which is an expectation of a non-linear expression.

Remark 8.8 *(*Interchanging the order of* E *and* ∂*) Consider expected utility in (8.1) and let the outcomes be functions of a portfolio weight* v*, as in* $x_s(v)$*. Differentiating wrt.* v *then gives*

$$\frac{d\,\mathrm{E}\,U(x)}{dv} = \sum_{s=1}^{S} \pi_s \frac{dU(x_s)}{dx}\frac{dx_s}{dv} = \mathrm{E}\,\frac{dU(x)}{dv},$$

where the last expression uses a short hand notation for how utility depends on v*.*

Clearly, the first order condition (8.7) defines one equation in one unknown (v). Unfortunately, it can be complicated. The expectation requires integration and marginal utility might be non-linear, together requiring numerical methods. Explicit solutions are only possible in a few simple cases.

Example 8.9 *(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 risk-free 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*

$$\max_v \pi \ln(vR^{e-} + R_f + 1) + (1-\pi)\ln(vR^{e+} + R_f + 1).$$

The first order condition ($\partial\,\mathrm{E}\,U(R_p)/\partial v = 0$*) is*

$$0 = \pi\frac{R^{e-}}{vR^{e-} + R_f + 1} + (1-\pi)\frac{R^{e+}}{vR^{e+} + R_f + 1},\ \text{so}$$

$$v = -(1+R_f)\frac{\pi R^{e-} + (1-\pi)R^{e+}}{R^{e-}R^{e+}}.$$

See Figure 8.3 for an illustration. As a special case, consider $R_f = 0$ *and* $R^{e-} = -1$*, so the bad state means losing the entire investment (bet). In this case,* $v = (1-\pi) - \pi/R^{e+}$*. This is often used to illustrate betting/log utility (the "Kelly" criterion).*

Remark 8.10 *(*When to put all investments in the risk-free asset?*) Suppose* $v = 0$ *would be an optimal decision, then the portfolio return equals the risk-free rate which is not random. The first order condition (8.7) can then be written*

$$\mathrm{E}[U'(R_f)R^e] = U'(R_f)\,\mathrm{E}\,R^e = 0$$

which holds only if $\mathrm{E}\,R^e = 0$*. This shows that it is optimal to make zero investment in the risky asset when its expected excess return is zero, which is intuitively reasonable.*

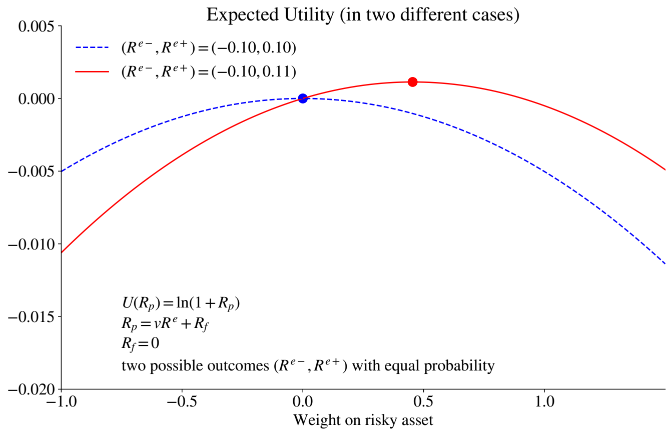


Figure 8.3: Example of portfolio choice with a log utility function

8.2.2 Utility-Based Portfolio Choice with Several Risky Assets

We now consider the case with n risky assets and a risk-free asset. The optimization problem is

$$\max_{v_1, v_2, \ldots} \mathrm{E}\, U(R_p), \text{ where} \tag{8.9}$$

$$R_p = \textstyle\sum_{i=1}^{n} v_i R_i^e + R_f. \tag{8.10}$$

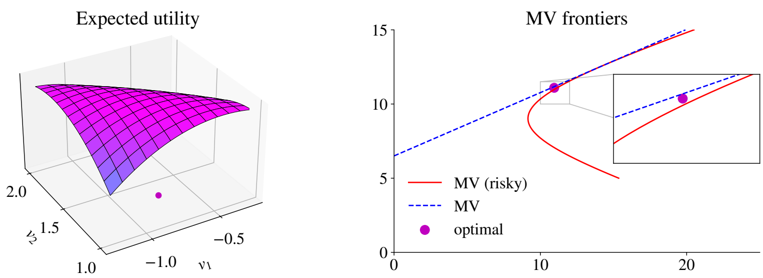
The first order conditions for the portfolio weights on the risky assets are

$$\mathrm{E}[U'(R_p) R_i^e] = 0 \text{ for } i = 1, 2, \ldots, n. \tag{8.11}$$

This is similar to the case with one risky asset, but now there are n (non-linear) equations in n unknowns: $v_1, v_2, \ldots, v_n$. For instance, with a CRRA utility function we get

$$\mathrm{E}\, \frac{R_i^e}{(\sum_{i=1}^{n} v_i R_i^e + R_f)^{\gamma}} = 0 \text{ for } i = 1, 2, \ldots, n. \tag{8.12}$$

Notice that calculating the expectation involves integrating over n dimensions. See Figures 8.4 –8.6 for illustrations. The (explicit or numerical) solution is often hard to obtain—so it would be convenient if we could simplify the problem.



Utility function: $(1+R)^{1-\gamma}/(1-\gamma), \gamma = 5$

Two risky assets (1 and 2) and one risk-free asset

Three states with equal probability, returns in %:

	state 1	state 2	state 3
asset 1	-3.0	8.0	20.0
asset 2	-4.0	22.0	15.0
risk-free	6.5	6.5	6.5

Portfolio weights:

	optimal	MV
asset 1	-0.73	-0.34
asset 2	1.32	0.94
risk-free	0.41	0.41

Figure 8.4: Example of when the optimal portfolio is (slightly) off the MV frontier

8.2.3 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 (8.11). In particular, sometimes we can show that the portfolio will be on the mean-variance frontier.

The optimal portfolio is on the mean-variance frontier when optimisation problem can be rewritten as a function in terms of the expected return (positive derivative) and the variance (negative derivative) only

$$\max_v V(\mathrm{E}\, R_p, \mathrm{Var}(R_p)), \tag{8.13}$$

where $\partial V()/\partial \, \mathrm{E}\, R_p > 0$ and $\partial V()/\partial \, \mathrm{Var}(R_p) < 0$. In this case, we should interpret $V()$ as incorporating the preferences, all relevant restrictions and also the features of the return distribution. See Danthine and Donaldson (2005) 4–6 and Huang and Litzenberger (1988) 4–5 for more detailed discussions.

This means that Figure 8.5 summarizes the preferences and the possibility set (everything below the CML).

In contrast, Figures 8.4 and 8.6 show examples when (8.13) does not hold. For instance, the preferences may include concerns about the skewness of the portfolio returns, at the same time as the return distribution exhibits non-trivial skewness. In such cases, the optimal portfolio may be off the MV frontier.

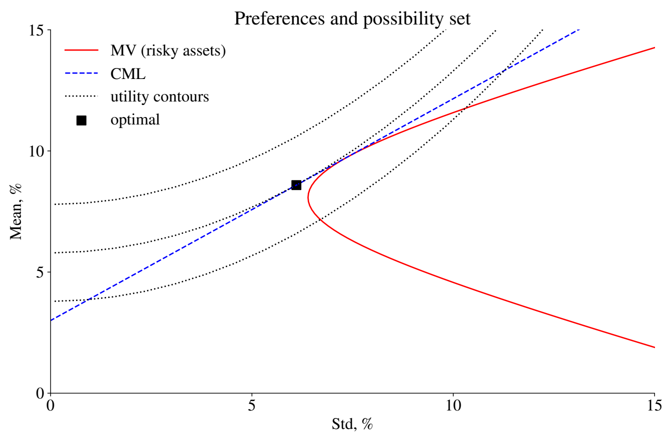


Figure 8.5: Iso-utility curves. The calculations use the properties of the assets in Table 8.1.

	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 8.1: Characteristics of the three assets in some 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, $\times 100^2$).

8.2.4 Special Cases

This section outlines special cases when the utility-based portfolio choice problem can be rewritten in terms of mean and variance only as in (8.13), so that the optimal portfolio is on the mean-variance frontier.

Case 1: Mean-Variance Utility

We already 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.)

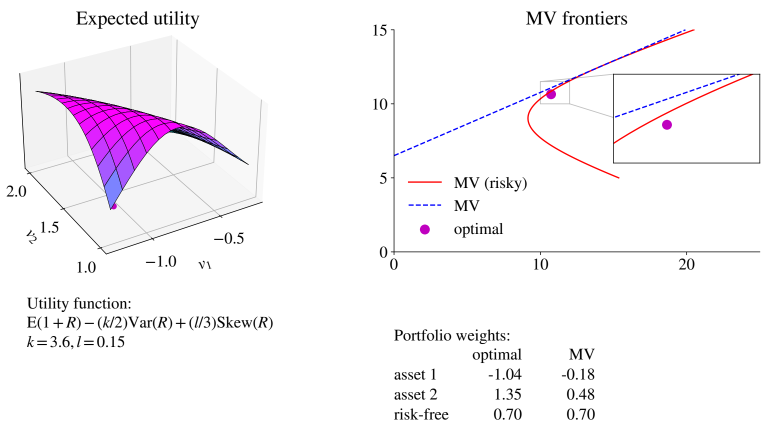


Figure 8.6: Example of when the optimal portfolio is (slightly) off the MV frontier

Case 2: Quadratic Utility

If utility is quadratic in the return (or equivalently, in wealth)

$$U(R_p) = R_p - kR_p^2/2, \tag{8.14}$$

then expected utility can be written

$$\mathrm{E}\, U(R_p) = \mathrm{E}\, R_p - k[\mathrm{Var}(R_p) + (\mathrm{E}\, R_p)^2]/2 \tag{8.15}$$

since $\mathrm{Var}(R_p) = \mathrm{E}\, R_p^2 - (\mathrm{E}\, R_p)^2$. 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 amount invested in risky assets decreases as wealth increases (increasing absolute risk aversion).

Case 3: Normally Distributed Returns

When the distribution (as perceived by the investor) of the investable assets is jointly normal, then all portfolio returns are normally distributed. In this case, maximizing $\mathrm{E}\, U(R_p)$ will result in a mean variance portfolio, at least if the utility function is strictly increasing and concave. The reason is that the expected value of such a utility function is increasing in the mean and decreasing in the volatility, and that these two moments fully describe a normal distribution.

Proposition 8.11 *If the returns of all investable assets are jointly normally distributed and the utility function is strinctly increasing and concave, maximizing* $\mathrm{E}\, U(R_p)$ *will result in a mean variance portfolio.*

Actually, the same result holds for any elliptical distribution with finite second moments (for instance, a multivariate t-distribution with more than 2 degrees of freedom), see Owen and Rabinovitch (1983).

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. *Third*, some assets with non-linear payoffs, like options, have return distributions that must be non-normal.

As an example of what happens when we combine a normal distribution with a valid utility function, consider the next propositions. Further examples and applications, for instance, using the Telser criterion, are discussed in a separate section below.

Proposition 8.12 *If returns are normally distributed, then maximizing the expected value a utility function with* constant absolute risk aversion *(CARA)* $k > 0$*,* $U(R_p) = -\exp(-R_p k)$*, is the same as solving a mean-variance problem. (The proof is in the appendix.)*

Case 4: CRRA Utility and Lognormally Distributed Portfolio Returns

Proposition 8.13 *Consider a CRRA utility function,* $(1+R_p)^{1-\gamma}/(1-\gamma)$*, and suppose all log portfolio returns,* $r_p = \ln(1+R_p)$*, are normally distributed. The solution is then, once again, on the mean-variance frontier. (The proof is in the appendix.)*

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.

Proof (of Proposition 8.11) First, joint normality of all returns means that portfolio returns are normally distributed. Second, a normal distribution $N(\mu_p, \sigma_p^2)$ is fully described by the mean and variance. Third (following Ingersoll (1987)), write $R_p \sim N(\mu_p, \sigma_p^2)$ as $\mu_p + \sigma_p z$ where $z \sim N(0, 1)$. Expected utility is then

$$\mathrm{E}\, U(R_p) = \mathrm{E}\, U(\mu_p + \sigma_p z).$$

The derivative with respect to μ_p is

$$\partial\, \mathrm{E}\, U(\mu_p + \sigma_p z)/\partial \mu_p = \mathrm{E}\, U'(\mu_p + \sigma_p z),$$

which is positive since $U'() > 0$. Also, the derivative with respect to σ_p is

$$\partial\, \mathrm{E}\, U(\mu_p + \sigma_p z)/\partial \sigma_p = \mathrm{E}[U'(\mu_p + \sigma_p z) z]. \qquad (*)$$

This must be negative since z has a symmetric distribution around zero, so for every term where $z = x$, there is also a term with $z = -x$. The expectation can thus be calculated as $\int_0^\infty [U'(\mu_p + \sigma_p x) - U'(\mu_p - \sigma_p x)] x \phi(x) dx$, where $\phi(x)$ is the $N(0, 1)$ pdf. The term in square brackets is negative since marginal utility is descreasing, that is, utility is concave ($U''() < 0$). For an alternative proof, using a Taylor series expansion of $\mathrm{E}\, U()$, see Pennacchi (2008) 2. □

8.2.5 Application of Normal Returns

This section gives a few examples of how fairly non-standard preferences, combined with normally distributed portfolio returns, give optimal portfolios on the mean-variance frontier.

The down-side risk measure Value at Risk (VaR) is just a quantile of the loss distribution, while Expected Shortfall (ES) is the average loss in case the loss is beyond the VaR. Target semivariance (TSV) is the average squared deviation around a target, but only counting the downside. Another chapter discusses the details and shows that, when returns are normally distributed, then *all three measures are increasing in the standard deviation*. Remark 8.14 summarises the key features, and details are in another chapter.

Remark 8.14 *(VaR, ES and TSV with normally distributed returns) If the return is normally distributed, $R \sim N(\mu, \sigma^2)$, then $VaR_\alpha = -(\mu + c\sigma)$, where c is the $1 - \alpha$ quantile of a $N(0, 1)$ distribution (for unstance, -1.64 for 5%). Also, $ES_\alpha = -[\mu - \phi(c)\sigma/(1 - \alpha)]$, where $\phi()$ is the pdf or a $N(0, 1)$ variable. Finally, it can be shown that the TSV $\lambda(h)$ is a strictly increasing function of the standard* *deviation,* $d\lambda_p(h)/d\sigma = 2\sigma\Phi(a)$*, where* $\Phi()$ *is the distribution function of a standard normal and* $a = (h - \mu)/\sigma$.

With normally distributed returns, the VaR, ES and TSV are strictly increasing functions of the variance. In this case, the portfolio that *minimizes the VaR, ES* or *TS* at a given average return will be on the mean-variance frontier.

Another portfolio choice approach is to *use the value at risk (VaR) as a restriction*. For instance, the *Telser criterion* maximizes the expected portfolio return subject to the restriction that the value at risk does not exceed a given level V^*

$$\max_v \mathrm{E}\, R_p \text{ st. } \mathrm{VaR}_\alpha < V^*. \tag{8.16}$$

When returns are normally distributed, Remark 8.14 shows that the restriction can be rewritten as

$$\mathrm{E}\, R_p > -V^* - c\, \mathrm{Std}(R_p), \tag{8.17}$$

where c is, for instance, -1.64 when the VaR_α has a confidence level $\alpha = 95\%$.

Example 8.15 *With a VaR confidence level of 95% and* $V^* = 0.1$*, then (8.17) gives* $\mathrm{E}\, R_p > -0.1 + 1.64\, \mathrm{Std}(R_p)$.

This optimization problem is illustrated in Figure 8.7. The objective is to find the portfolio with the highest expected return that satisfies the VaR restriction, which means that is has to be on or above the line defined by (8.17). Also, only points on or below the CLM (the mean-variance frontier based on both risky assets and a risk-free asset) are feasible.

The optimal portfolio is therefore where the restriction intersects the CLM: the Telser criterion applied to normally distributed returns gives a mean-variance portfolio. To be precise, the optimal portfolio puts

$$w = -\frac{R_f + V^*}{\mu_T^e + c\sigma_T}. \tag{8.18}$$

in the tangency portfolio (with average excess return μ_T^e and standard deviation σ_T) and the rest $(1 - w)$ in the risk-free asset.

We could instead use a restriction on expected shortfall or target semivariance, which define areas in a MV figure similar to that in Figure 8.7.

Example 8.16 *(Optimal portfolio. Telser) Let* $\mu_T^e = 6.7\%$*,* $\sigma_T = 7.4\%$ *and* $R_f =$

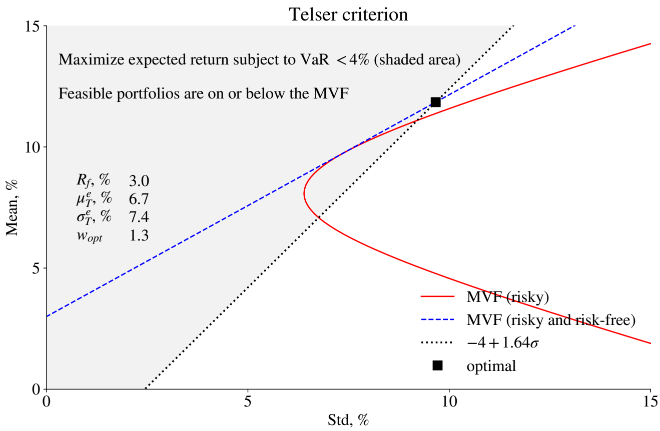


Figure 8.7: Telser criterion and VaR. The calculations use the inputs from Table 8.1.

3%. *The optimal portfolio with* $V^* = 4\%$ *is then*

$$w = -\frac{0.03 + 0.04}{0.067 - 1.64 \times 0.074} \approx 1.3.$$

Instead, if the restriction is that VaR$< 2\%$*, then the weight is* $w \approx 0.9$*. Figure 8.7.*

Proof of (8.18). As usual, the average return on a portfolio p of the CML is

$$\mu_p = R_f + SR_T \sigma_p,$$

where SR_T is the the Sharpe ratio of the tangency portfolio. This equals the mean return required by the VaR restriction (8.17) when

$$\sigma_p = -\frac{R_f + V^*}{SR_T + c}.$$

Since $\sigma_p = w\sigma_T$ (assuming $w \geq 0$), the optimal portfolio weight on the tangency portfolio is (8.18). □

8.3 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.

8.3.1 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 depends 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).

8.3.2 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.

8.3.3 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.

8.4 Appendix – Risk Aversion and the Level of Wealth*

This section discusses how risk aversion is related to the wealth level. (Notice that when we use the portfolio return as the argument of the utility function, then this amounts to disregarding differences across wealth levels.)

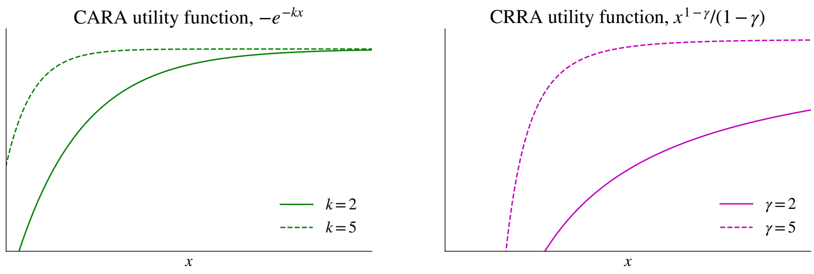


Figure 8.8: Examples of utility functions

First, define *absolute risk aversion* as

$$A(W) = \frac{-U''(W)}{U'(W)}, \tag{8.19}$$

where $U'(W)$ is the first derivative and $U''(W)$ the second derivative. Second, define *relative risk aversion* as

$$R(W) = WA(W) = \frac{-WU''(W)}{U'(W)}. \tag{8.20}$$

These two definitions are strongly related to the attitude towards taking risk (see below).

Figure 8.8 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/risk-free trade-off) is constant across wealth levels. This means, of course, that wealthy investors would have a lower portfolio weight on risky assets.

Remark 8.17 *(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 wealth level. This fits with historical data which show no trends in portfolio weights or risk premia—in spite of investors having become much richer over time.

Remark 8.18 *(Risk aversion in CRRA utility function)* $U(W) = W^{1-\gamma}/(1-\gamma)$ *gives* $U'(W) = W^{-\gamma}$ *and* $U''(W) = -\gamma W^{-\gamma-1}$*, so we have* $A(W) = \gamma/W$ *and* $R(W) = \gamma$*. 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. The investor is indifferent if

$$\mathrm{E}\, U(W+Z) = U(W-P). \tag{8.21}$$

If Z is a small risk, then we can use a second order approximation and solve for the price as

$$P \approx A(W)\,\mathrm{Var}(Z)/2. \tag{8.22}$$

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 8.19 *(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 (8.22). First, 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 $\mathrm{E}\, Z = 0$. Second, approximate $U(W-P) \approx U(W) - U'(W)P$. Finally, make the two approximations equal to get (8.22). □

If we change the setting in (8.21)–(8.22) to make the risk proportional to wealth, that is $Z = Wz$ where z is the risk factor, then (8.22) directly gives

$$\begin{aligned}P &\approx A(W)W^2\,\mathrm{Var}(z)/2, \text{ so}\\ P/W &\approx R(W)\,\mathrm{Var}(z)/2.\end{aligned} \tag{8.23}$$

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 8.20 *(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 in risky assets, whereas a high relative risk aversion typically leads to small *portfolio weights* of risky assets.

8.5 Appendix – Portfolio Choice with $N()$ Returns*

8.5.1 Case 3 and 4: Proofs

Proof of Proposition 8.12. First, recall that if $x \sim N(\mu, \sigma^2)$, then $\mathrm{E}\, e^x = e^{\mu+\sigma^2/2}$. Therefore, rewrite expected utility as

$$\mathrm{E}\, U(R_p) = \mathrm{E}[-\exp\left(-R_p k\right)] = -\exp[-\mathrm{E}\, R_p k + \text{Var}(R_p)k^2/2].$$

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

$$-\ln(-z)/k = \mathrm{E}\, R_p - \text{Var}(R_p)k/2.$$

Therefore, maximizing the expected CARA utility or MV preferences (in terms of the returns) gives the same solution. □

Proof of Proposition 8.13. Notice that

$$\mathrm{E}(1 + R_p)^{1-\gamma}/(1-\gamma) = \mathrm{E}\exp[(1-\gamma)r_p]/(1-\gamma), \text{ where } r_p = \ln(1 + R_p).$$

Since r_p is normally distributed, the expectation is (recall that if $x \sim N(\mu, \sigma^2)$, $\mathrm{E}\exp(x) = \exp(\mu + \sigma^2/2)$)

$$\mathrm{E}\exp[(1-\gamma)r_p]/(1-\gamma) = \exp[(1-\gamma)\,\mathrm{E}\, r_p + (1-\gamma)^2\, \text{Var}(r_p)/2]/(1-\gamma).$$

If $\gamma > 1$ ($0 < \gamma < 1$), then the function $\ln[z(1-\gamma)]/(1-\gamma)$ is defined for $z < 0$ ($z > 0$) and it is increasing in z. Let z be the right hand side of the previous equation (which is negative if $\gamma > 1$ and positive if $0 < \gamma < 1$) and apply the transformation to get

$$\mathrm{E}\, r_p + (1-\gamma)\, \text{Var}(r_p)/2.$$

To express this in terms of the mean and variance of the return instead of the log return we use the following facts: if $\ln y \sim N(\mu, \sigma^2)$, then $\mathrm{E}\, y = \exp(\mu + \sigma^2/2)$

and $\operatorname{Std}(y) / \operatorname{E} y = (\exp(\sigma^2) - 1)^{1/2}$. Using this fact in the previous expression and rearranging (express $\operatorname{E} r_p$ and $\operatorname{Var}(r_p)$ in terms of $\operatorname{E} R_p$ and $\operatorname{Var}(R_p)$) gives

$$\ln(1 + \operatorname{E} R_p) - \gamma \ln[\operatorname{Var}(R_p)/(1 + \operatorname{E} R_p)^2 + 1]/2,$$

which is increasing in $\operatorname{E} R_p$ and decreasing in $\operatorname{Var}(R_p)$. We therefore get a mean-variance portfolio. $\square$

Chapter 9

Multi-Factor Models

Earlier chapters have discussed both theoretical and empirical problems with CAPM. This chapter will therefore discuss multi-factor models. It provides theoretical motivations of some multi-factor models and performs an empirical test of a commonly used model (Fama and French (1993)).

Theoretical arguments about multi-factor models can easily become somewhat cumbersome, but the proofs can be skipped over at first reading.

9.1 Factor Investment

A number of factor returns related to, for instance, firm characteristics like size and profitability, have shown good performance over long periods of time. It is therefore common to base investment strategies on those characteristics—and a large number of funds and other investment vehicles have been developed for this purpose. This approach is called *factor investing* or "smart beta." It is essentially a dynamic trading strategy since firm characteristics change over time.

In studies of investment fund performance, it is often found that the abnormal performance (α from a CAPM regression) can be explained by a fairly small set of factors. It seems that many fund managers have indeed been able to invest in those characteristics that have historically paid off. This suggests that the market might have moved beyond CAPM and that a multi-factor model would be empirically more appropriate.

Empirical Example 9.1 *(Fama-French factors) Figure 9.1 illustrate several of the factors discussed by Fama and French (1993) and Fama and French (2015), while Table 9.1 summarises the return patterns. Many factors have positive excess returns,*

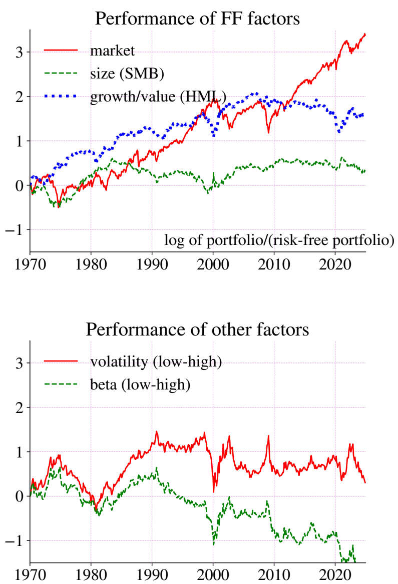


Figure 9.1: Cumulated returns of important equity factors

which is especially interesting since they are long-short in equities. It is also clear that several of the factors are virtually uncorrelated with the market excess return, so the α values are considerable (and similar to the average excess returns).

9.1.1 Portfolio Sorts

Many of the factors in Table 9.1 are based on portfolio sorts, which can be thought of as dynamic trading strategies. This approach is often used to study how asset characteristics relate to returns, sometimes as an alternative to regressions.

In a *univariate sort*, we could rank (sort) the assets according to z_{it} and then construct portfolios (say, three equally weighted portfolios: low, mid and high). Then, we measure and study the returns of the portfolios. Typically, the sorting is repeated at regular time intervals (every day or perhaps every June), so z_{it} should be understood as the value used in a particular period.

Empirical Example 9.2 *(Sorting on recent returns) See Table 9.2 for an empirical example where the 25 FF portfolios are sorted into low/low recent 22-day returns, with 5 portfolios in each. The results indicate strong momentum.*

	μ, %	σ, %	SR	β	α, %
market	7.30	15.92	0.46	1.00	−0.00
size	1.18	10.66	0.11	0.19	−0.20
growth/value	3.44	10.73	0.32	−0.14	4.49
momentum	6.94	14.99	0.46	−0.18	8.24
profitability	3.68	7.89	0.47	−0.10	4.41
investment	3.51	7.18	0.49	−0.17	4.72
volatility (low-high)	3.26	22.70	0.14	−0.88	9.69
beta (low-high)	−2.26	20.41	−0.11	−0.83	3.83

Table 9.1: Descriptive statistics of excess returns of different US equity portfolios (including the Fama-French factors and more), annualised. Monthly data 1970:01-2024:12.

	Portfolio returns
Low 22-day return	4.15
	(0.20)
	[−5.36]
High 22-day return	13.71
	(0.74)
	[5.16]
Difference (H-L)	9.55
	(0.85)
	[10.52]

Table 9.2: Average excess returns, (Sharpe ratios) and [α] for 3 portfolios from a univariate sort on recent (22-day) returns (5/5 assets). Annualized figures. Daily data on 25 FF portfolios 1979:01-2024:12

Bivariate (double) sorts are used when there are two important characteristics (here called x and z) and you want to study how z affects returns, holding x "constant". This may be important if x and z are correlated. An *independent bivariate sort* first does a univariate sort based on x_{it} (say, forming 3 categories: growth, neutral or value), then it makes another univariate sort according to the other sorting variable z_{it} (say, forming two categories: small or big). Then, we find the intersections of the two sorts. In contrast, in a *dependent bivariate sort* we first sort according to x_{it} as before. Then, *within* an x_t category we sort according to z_{it}. This allows us to control the number of assets in each group.

9.2 An Overview of Multi-Factor Models

This section gives a short introduction to multi-factor models. Model details and proofs are in later sections.

A multi-factor model extends the market model by allowing more factors to explain the return on an asset. For instance, a two-factor model is

$$R_{it}^e = \alpha + \beta_{im} R_{mt}^e + \beta_{ic} R_{ct}^e + \varepsilon_{it}, \tag{9.1}$$

where R_m^e is the excess return on the market and R_c^e is the excess return on some other portfolio. As usual, we require $\mathrm{E}\,\varepsilon_{it} = 0$, and that ε_{it} is uncorrelated to all regressors.

The pricing implication is a multi-beta model

$$\mathrm{E}\,R_i^e = \beta_{im}\mu_m^e + \beta_{ic}\mu_c^e. \tag{9.2}$$

Notice that there is no intercept, so α in (9.1) should be zero.

Remark 9.3 *(When factors are not excess returns*) Equation (9.2) 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 (9.1) and (9.2). Second, we could instead* (a) *estimate the betas (β_{im}, β_{ic}) by a time series regression of (9.1), but allowing for an intercept; and (b) estimate the factor risk premia (μ_m^e, μ_c^e) by a cross-section of (9.2) where the dependent variable is the historical average returns of different assets ($i = 1, 2, ...n$) and the regressors are the betas from the first step.*

This chapter will discuss *theoretical* multi-factor models: *(a)* CAPM with background risk as well as *(b)* a consumption-based model.

There are also many *empirically motivated* multi-factor models that have been found to work well in practice. 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 empirically performs much better than CAPM. The more recent Fama and French (2015) extends this to a five-factor model. Also, the multi-factor model by MSCI Barra (MSCI Inc. (2024)) is widely used in the financial industry. It uses a set of firm characteristics as factors, for instance, size, volatility, price momentum, and industry/country (see Stefek (2002)). These models suggest that CAPM may not be enough.

More detailed treatments of this important topic is found in, for instance, Cochrane (2005) and Back (2010).

9.3 Portfolio Choice with Background Risk

This section discusses the portfolio problem when there is "background risk" or non-traded assets (see Mayers (1972)), for instance, labour income, real estate, a private business, or a liability.

The existence of background risk/non-traded assets will affect portfolio choice, and therefore perhaps also asset prices.

9.3.1 Portfolio Choice with Background Risk: One Risky Asset

Consider a mean-variance investor who forms a financial portfolio by choosing between a risky asset (henceforth called "equity") with return R and a risk-free asset with return R_f. The investor also has a background risk in the form of an endowment (positive or negative) of a non-traded asset with return R_c. This could, for instance, be labour income or a house (positive endowment) or a liability (negative endowment).

The non-traded asset accounts for the fraction ϕ of total wealth, while the financial portfolio accounts for $1-\phi$. When the non-traded asset is a liability, then $\phi < 0$. The return on the total portfolio, R_p, is

$$R_p = (1-\phi)R_{Fin} + \phi R_c, \text{ with} \tag{9.3}$$

$$R_{Fin} = wR^e + R_f, \tag{9.4}$$

where R_c is the return (change of value, payoff) of the non-traded asset.

The investor chooses w to maximize

$$\mathrm{E}\, U(R_p) = \mathrm{E}\, R_p - \frac{k}{2}\,\mathrm{Var}(R_p), \tag{9.5}$$

and the optimal value is

$$w = \frac{\mu^e/k - \phi S_c}{(1-\phi)\sigma^2}, \tag{9.6}$$

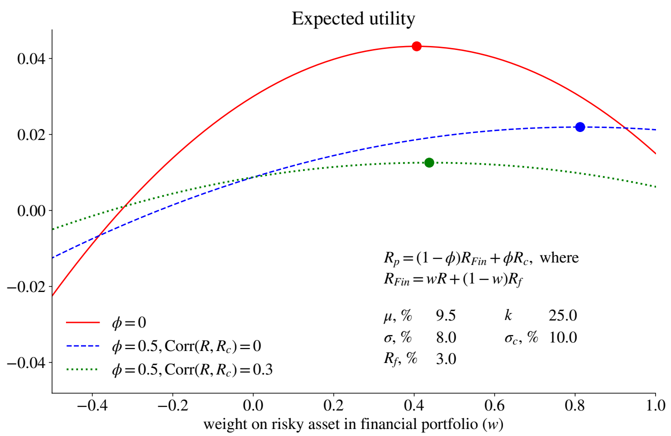


Figure 9.2: Portfolio choice with non-traded assets

where σ^2 is the variance of equity and S_c is the covariance of equity and the non-traded asset.

Proof of (9.6). To simplify the notation, write the portfolio return as $R_p = vR^e + \phi R_c^e + R_f$, where $v = (1-\phi)w$. Use in the objective function to get

$$\mathrm{E}\, U(R_p) = v\mu^e + \phi\mu_c^e + R_f - \frac{k}{2}(v^2\sigma^2 + \phi^2\sigma_c^2 + 2v\phi S_c),$$

The first order condition wrt. v is $\mu^e - k(v\sigma^2 + \phi S_c) = 0$, solve for v and divide by $1-\phi$ to get (9.6). □

The second term in the optimal portfolio weight, *the hedging term*, depends on how important the non-traded asset is in the portfolio (ϕ) and also on the covariance term (S_c). Clearly, if there is no non-traded asset in the total portfolio ($\phi = 0$), then we are back in a traditional MV case.

Remark 9.4 *(Interpreting the hedging term*) The hedging term in (9.6) is related to the slope coefficient in a regression of the non-traded asset on equity, $R_{ct}^e = \alpha + \gamma R_{it}^e + \varepsilon_t$, since $\gamma = \sigma_{ic}/\sigma_i^2$, which corresponds to S_c/σ^2 in the notation above.*

Several things can be noticed. First, *when the correlation is zero* (Corr$(R, R_c) = 0$), then the equity weight is increasing in the amount of non-traded assets (ϕ), while the opposite holds for the risk-free asset; see Figure 9.2 for an illustration. The intuition is that a zero correlation means that the non-traded asset 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 towards equity.

Second, *when the correlation is positive* ($\mathrm{Corr}(R, R_c) > 0$) and we have a positive exposure to the non-traded asset ($\phi > 0$), then the hedging term will reduce the equity weight. Again, see Figure 9.2. The intuition is that the overall portfolio now includes a lot of "equity like" assets, so the financial portfolio should be tilted towards the risk-free asset. The opposite holds when the exposure to the non-traded asset is negative (a liability, $\phi < 0$) or when the non-traded asset is negatively correlated with equity.

Example 9.5 *(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. *In this case, the non-traded asset 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 this section suggests that a low correlation of stock returns and wages means that the young investor is endowed with a bond-like asset, so the financial portfolio should be tilted towards equity. Old investors less so.*

9.3.2 Portfolio Choice with Background Risk: Several Risky Assets

With several risky assets, the financial portfolio return is

$$R_{Fin} = w' R^e + R_f, \tag{9.7}$$

where w now is a vector of portfolio weights, R a vector of returns on the risky assets and $\mathbf{1}$ is a vector of ones. The optimal portfolio is now

$$w = \Sigma^{-1} \frac{\mu^e/k - \phi S_c}{1 - \phi}, \tag{9.8}$$

where Σ is the variance-covariance matrix of the risky assets (not including the non-traded asset) and S_c is a vector of covariances of the assets with the non-traded asset. The portfolio weights of the financial subportfolio will (as long as $\phi S_c \neq 0$) give a return that is *off the mean-variance frontier*—and will differ across investors if the non-traded asset does: the *two-fund separation theorem is no longer valid*. See Figure 9.3 for an illustration. Note that the optimal portfolio tends to have lower weights on assets that are positively correlated with the non-traded asset and vice versa (compare with Table 9.3).

	μ, %	Σ, bp				$\rho_{x,c}$
		A	B	C	c	
A	11.5	166	34	58	161	0.50
B	9.5	34	64	4	180	0.90
C	6.0	58	4	100	-25	–0.10
c	10.0	161	180	-25	625	1.00

Table 9.3: Characteristics of the assets in the example of MV with background risk. Notice that μ, % is the expected return in % (that is, $\times 100$) and Σ, bp is the covariance matrix in basis points (that is, $\times 100^2$). $\rho_{x,c}$ are the correlations of each assset with the background risk.

Proof of (9.8). The portfolio return (9.7) can be written $R_p = v'R^e + \phi R_c^e + R_f$, where $v = (1-\phi)w$. The investor solves

$$\max_v v'\mu^e + \phi\mu_c^e + R_f - \frac{k}{2}(v'\Sigma v + \phi^2\sigma_c^2 + 2\phi v' S_c),$$

with first order conditions

$$\mu^e - k(\Sigma v + \phi S_c) = 0.$$

Solve for v and divide by $1-\phi$ to get (9.8). □

Remark 9.6 *(Interpreting the hedging terms*) The hedging terms are related to the slope coefficients from a regression of R_c^e on the vector of investable risky assets (R^e) $R_{ct}^e = \alpha + \gamma' R_t^e + \varepsilon_t$, since $\gamma = \Sigma^{-1} S_c$.*

Example 9.7 *(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 more modest. This person should therefore tilt the financial portfolio away from pharmaceutical stocks: the market portfolio is not the best for everyone.*

Remark 9.8 *(Transformed assets*) However, the optimal portfolio w is on the mean-variance frontier of some transformed assets with returns Z_i. We can rewrite the portfolio return as*

$$R_p = w'Z + (1 - \mathbf{1}'w)Z_f, \text{ where}$$
$$Z_i = (1-\phi)R_i + \phi R_c \text{ and and } Z_f = (1-\phi)R_f + \phi R_c.$$

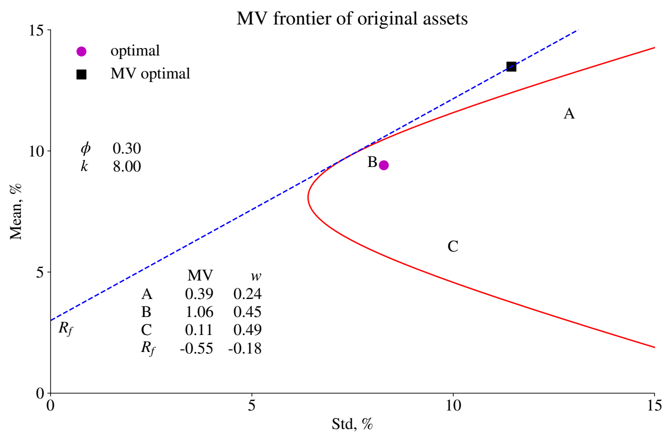


Figure 9.3: Portfolio choice with background risk. The properties of the assets are shown in Table 9.3.

Notice that all these transformed assets (also Z_f) are risky. The optimal portfolio will be on the mean-variance frontier of $Z = (Z_i, Z_f)$. See Figure 9.4. (The "proof" is that maximizing the objective function (9.5) subject to this new definition of the portfolio return is a traditional mean-variance problem—but in terms of the transformed assets, Z.)

9.4 Asset Pricing Implications

Background risk that varies greatly across investors is unlikely to affect asset prices. Hedging it is then rather similar to hedging idiosyncratic (asset specific) risk.

In contrast, if the background risk affects the portfolio choice of a *large fraction* of the investors, then it is also likely to influence asset prices. For instance, an asset that provides an effective hedge against a common background risk will be greatly demanded—and therefore generate low returns.

To create tractable pricing expressions, we use a *factor mimicking portfolio* λ (or factor replicating portfolio, see Cochrane (2005) 6 and Back (2010) 6) in place of the true factor (here, the background risk). This is the portfolio with the highest squared correlation (R^2) with the true factor. In fact, it is the fitted value (minus the intercept) from a linear regression of the background risk on all excess returns. In practice, an approximate factor mimicking portfolio might used to facilitate the

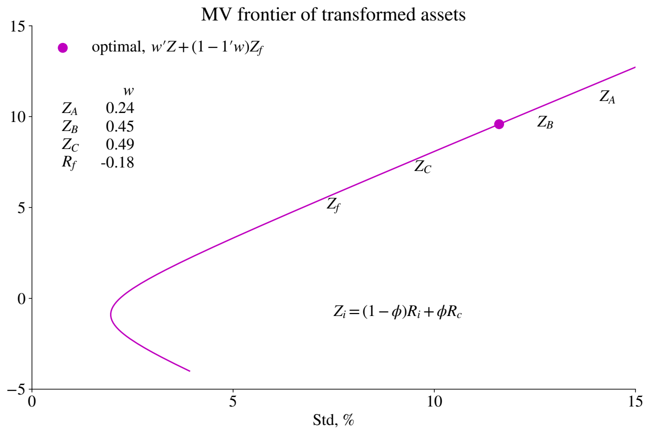


Figure 9.4: Portfolio choice with background risk, transformed assets. The properties of the assets are shown in Table 9.3.

implementation.

Notice that a factor mimicking portfolio is not a new asset, so it does not change any prices. Rather, it is just a convenient way of characterizing the pricing process.

The market portfolio m is here chosen to be the optimal portfolio for an investor with a risk aversion (k) such that the portfofolio weight on the risk-free asset is zero. This, together with the factor mimicking portfolio will define a two-factor model.

Remark 9.9 *(Factor mimicking portfolio) Regress $R_c^e = \delta + R^{e\prime} w_\lambda + \varepsilon$ and notice that $w_\lambda = \Sigma^{-1} S_c$, where Σ is the variance-covariance matrix of the assets and S_c the vector of covariances of the assets with the background risk. The factor mimicking portfolio has the excess return $R_\lambda^e = w_\lambda' R^e$. It is clear that, for any portfolio w_p, $\sigma_{pc} = \sigma_{p\lambda}$, since ε is uncorrelated with R^e.*

Example 9.10 *(Numerical example I) Table 9.4 shows results based on the asset properties in Table 9.3: the portfolio weights on the risky assets for the factor mimicking portfolio (λ), the market portfolio (m), and a randomly picked portfolio (p) that will be used to test the pricing ability of the two-factor model below.*

The theoretical implication of the background risk is a *multi-beta model*

$$\mathrm{E}\, R_p^e = \beta_{pm}\mu_m^e + \beta_{p\lambda}\mu_\lambda^e, \tag{9.9}$$

	w_m	w_λ	w_p
A	0.19	0.74	0.50
B	0.33	2.47	0.40
C	0.48	−0.78	0.10

Table 9.4: Portfolio weights on the market portfolio (w_m), the factor mimicking portfolio (w_λ), and a randomly picked portfolio w_p. The remainder is invested in the riks-free asset. The market portfolio is the optimal portfolio when $k = 8.691$. Based on the asset properties in Table 9.3.

where μ_m^e and μ_λ^e are the average excess returns on the two factors, and the betas are from the linear regression

$$R_p^e = \alpha + \beta_{pm} R_m^e + \beta_{p\lambda} R_\lambda^e + \varepsilon, \tag{9.10}$$

where we regress the excess return of a portfolio p on the excess returns on the two factors.

Clearly, the expected excess return on portfolio p in (9.9) depends on how it is related to both the (financial) market and the background risk. Notice that there is no intercept in (9.9), so the testable implication is that $\alpha = 0$ in (9.10). (The formal proof is in the Appendix.)

Example 9.11 *(Numerical example II) Table 9.5 shows the implied expected excess returns of (m, λ), their variance-covariance matrix, and also their covariances with portfolio p. The betas, shown in the last column, are for the regression. According to these results, the expected excess return of portfolio p should be*

$$\mathrm{E}\, R_p^e = 0.8 \times 5.22\% + 0.15 \times 19.98\% \approx 7.17\%.$$

The actual value is 7.15% (see the caption of the table). The difference is due to rounding.

Remark 9.12 *(Calculating the properties of $(m, \lambda, p)^*$) With the portfolio weights, we can calculate the variance-covariance matrix of (m, λ) and also the vector of covariance of p with (m, λ) as*

$$\begin{bmatrix} \sigma_{mm} & \sigma_{m\lambda} \\ \sigma_{m\lambda} & \sigma_{\lambda\lambda} \end{bmatrix} = \begin{bmatrix} w_m' \Sigma w_m & w_m' \Sigma w_\lambda \\ w_m' \Sigma w_\lambda & w_\lambda' \Sigma w_\lambda \end{bmatrix} \text{ and } \begin{bmatrix} \sigma_{pm} \\ \sigma_{p\lambda} \end{bmatrix} = \begin{bmatrix} w_p' \Sigma w_m \\ w_p' \Sigma w_\lambda \end{bmatrix}$$

	μ^e, %	Cov($[m,\lambda]$), bp		Cov(x,p), bp	β
m	5.22	51.98	78.81	53.23	0.80
λ	19.98	78.81	582.44	150.03	0.15

Table 9.5: Properties of the market (m), factor mimicking (κ), and the randomly selected portfolio (p). Expected returns are in percent and variance and covariances are in basis points. The β are from the regression $R_p^e = \alpha + \beta'[R_m^e, R_\lambda^e] + \varepsilon$. Also, $\mu_p^e = 7.15\%$. Based on the asset properties in Table 9.3 and results in Table 9.4.

Excess returns are calculated as $\mu_x^e = w_x'\mu^e$, where μ^e is the vector of excess returns for the investable assets. The slope coefficients in (9.10) are

$$\begin{bmatrix} \beta_{pm} \\ \beta_{p\lambda} \end{bmatrix} = \begin{bmatrix} \sigma_{mm} & \sigma_{m\lambda} \\ \sigma_{m\lambda} & \sigma_{\lambda\lambda} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{pm} \\ \sigma_{p\lambda} \end{bmatrix}.$$

9.4.1 Asset Pricing Implications II: Reinterpreting alpha

Consider the standard CAPM regression

$$R_{it}^e = \alpha_i + \beta_i R_{mt}^e + \varepsilon_{it}, \tag{9.11}$$

where R_{mt}^e is the market excess return in period t. We use time series data to estimate it. The intercept α is likely to be non-zero if the R_{it} returns are driven by a multi-factor model. That is, (9.11) suffers from an omitted variable bias.

To be precise, suppose the two-factor model (9.10) holds with a zero intercept. Then, the OLS estimate of α_i from (9.11) is

$$\hat{\alpha}_i = \hat{\theta}_0 \hat{\beta}_{ic}, \tag{9.12}$$

where $\hat{\beta}_{ic}$ is the beta in the two-factor model regression (9.10) and $\hat{\theta}_0$ is the estimate of the intercept in

$$R_{ct}^e = \theta_0 + \theta_1 R_{mt}^e + \eta_t. \tag{9.13}$$

(To show this, apply Remark 9.13.) Together, these two equations suggest that non-zero alphas from CAPM regression may be explained by a combination of *(1)* a missing factor ($\beta_{ic} \neq 0$); *(2)* and that factor is not "priced" by the market returns alone ($\theta_0 \neq 0$).

Remark 9.13 *(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 $y_t = x_t'\delta + \varepsilon_t$*

by OLS. It is well known that the OLS estimate is $\hat{\delta} = \hat{\beta} + \hat{\theta}\hat{\gamma}$, *where* $\hat{\theta}$ *is from regressing* $h_t = x_t'\theta + \eta_t$.

9.5 Joint Portfolio and Savings Choice

The basic *consumption-based* multi-period investment problem assumes 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 the task is to maximize expected utility

$$\max U(c_1) + \delta\,\mathrm{E}_1\,U(c_2), \text{ subject to} \tag{9.14}$$

$$c_1 + I_1 = W_1 \tag{9.15}$$

$$c_2 + I_2 = (1 + R_p)I_1 + y_2, \text{ where} \tag{9.16}$$

$$R_p = v'R^e + R_f. \tag{9.17}$$

In equation (9.14), 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 then. The constant δ is the time discounting, with $0 < \delta < 1$ indicating impatience. (In an equilibrium without risk, we will get a positive real interest rate if investors are impatient.)

Equation (9.15) is the budget constraint for period 1: an initial wealth (including exogenous income), W_1, is split between consumption, c_1, and investment, I_1. Equation (9.16) 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 (see (9.17)) depends on the portfolio weights v chosen in period 1 as well as on the returns on the assets.

Obtaining closed-form solutions is typically difficult. However, we can gain some insights by studying the first order conditions. The optimization problem involves maximizing with respect to the investment level (mostly a macro topic, but summarized in a Remark below) and how to form the investment portfolio, which is the focus here.

9.5.1 Optimal Portfolio Choice

This section studies the *portfolio choice*, that is, the portfolio weights in the vector v.

$$\mathrm{E}_1[U'(c_2)R_i^e] = 0 \text{ for } i = 1, \ldots, n. \tag{9.18}$$

The expression says that excess returns should be "orthogonal" to marginal utility. This is similar to earlier results on utility based portfolio choice, with the difference that marginal utility now depends on consumption rather than the portfolio return.

Proof (of (9.18)) The first order condition for v_i is $\delta\,\mathrm{E}_1[U'(c_2)\partial c_2/\partial v_i] = 0$. Differentiates the budget constraint (9.16) to get $\partial c_2/\partial v_i = I_1 R_i^e$ and simplify the resulting expression by dividing both sides by δI_1 (which is know in t and therefore can be moved outside the expectation). $\square$

The first order conditions (9.18) still contain some useful information, especially if we rewrite them as

$$\mathrm{E}_1\, R_i^e = -\operatorname{Cov}_1[U'(c_2), R_i^e]/\,\mathrm{E}_1\, U'(c_2), \tag{9.19}$$

where the time subscripts on the expectation and covariance operators indicate that they are conditional on the information in period 1.

Proof (of (9.19)) Recall that, by definition, $\operatorname{Cov}(x, y) = \mathrm{E}\, xy - \mathrm{E}\, x \times \mathrm{E}\, y$. $\mathrm{E}\, xy = 0$, so $\mathrm{E}\, y = -\operatorname{Cov}(x, y)/\,\mathrm{E}\, x$. Set $y = R_i^e$, $x = U'(c_2)$ and notice that (9.18) says that $\mathrm{E}\, xy = 0$. $\square$

First, the denominator is positive (marginal utility always is). Second, suppose the return is procyclical, $\operatorname{Cov}(c_2, R_i^e) > 0$. This will make $\operatorname{Cov}[U'(c_2), R_i^e] < 0$, since marginal utility $U'(c_2)$ is a decreasing function of the consumption level as the utility function is concave, see Figure 9.5. Together, this creates a positive risk premium, $\mathrm{E}\, R_i^e > 0$. That is, *an asset is risky if it is procyclical. (*Recall that risky assets have high risk premia since otherwise no one would like to buy those assets.)

Remark 9.14 *(*Linearizing $U'(c)$) A first-order Taylor approximation of marginal utility around $\bar{c}$ is $U'(C) \approx U'(\bar{c}) + U''(\bar{c})(c - \bar{c})$. The numerator in (9.19) can thus be written $-\operatorname{Cov}[U'(c_2), R_i^e] \approx -U''(\bar{c})\operatorname{Cov}(c_2, R_i^e)$,where $-U''(\bar{c}) > 0$ 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,

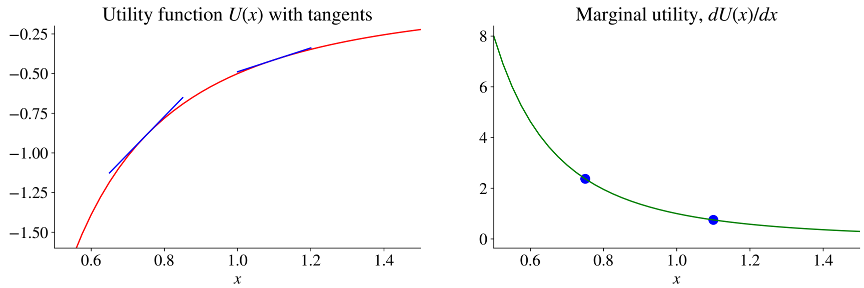


Figure 9.5: Utility and marginal utility

the objective function is

$$U(c_1) + \delta \, \mathrm{E}_1 \, U(c_2) + \delta^2 \, \mathrm{E}_1 \, U(c_3) + \ldots \delta^{T-1} \, \mathrm{E}_1 \, U(c_T). \tag{9.20}$$

Remark 9.15 *(**Investment level *in period 1, I_1) The first order condition for I_1 is that the derivative of (9.14) wrt I_1 is zero, $-U'(c_1) + \delta \, \mathrm{E}_1[U'(c_2)(1 + R_p)] = 0$, where $U'(c_t)$ is the marginal utility in period t. This says that consumption should be planned such 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. For instance, with logarithmic utility we get $\mathrm{E}_1 \, c_2/c_1 = \delta(1 + R_f)$, which says that when R_f is high, then the expected (planned) consumption path is upward sloping. (It is also clear that $I_2 = 0$ since investing in period 2 is just a waste.)*

9.5.2 The Stochastic Discount Factor or Pricing Kernel*

More advanced asset pricing theory often work with a *stochastic discount factor* (SDF) or pricing kernel. For the optimization problem (9.14)–(9.17) this could be

$$M_2 = \delta U'(c_2) \, / \, U'(c_1) \, . \tag{9.21}$$

Rewrite the first order condition for v_i (9.18) (notice that $\delta/U'(c_1)$ makes no difference since it is known at the time of investment) and the expression for expected excess returns (9.19) as

$$\mathrm{E}(MR_i^e) = 0, \text{ and} \tag{9.22}$$

$$\mathrm{E}\, R_i^e = -\mathrm{Cov}(M, R_i^e)/\, \mathrm{E}\, M, \tag{9.23}$$

where we (for convenience) drop all time subscripts. (Similarly, the foc in Remark 9.15 can be written $\mathrm{E}[M(1+R_p)] = 1$.) Notice that the *same* SDF is used for each asset i. Such SDFs can be derived in many ways (here we used a consumption plan approach), but they typically imply an expression like (9.22).

Remark 9.16 *(Pricing with a stochastic discount factor, SDF*) Let M_{t+1} be an SDF and x_{t+1} the payoff of an asset in $t+1$. Most asset pricing theories imply that that the price today of the asset today (P_t) must satisfy (a) $P_t = \mathrm{E}_t\, M_{t+1}x_{t+1}$. This implies that the gross return must satisfy (b) $\mathrm{E}_t\, M_{t+1}(1+R_{t+1}) = 1$ and the excess returns must satisfy (c) $\mathrm{E}_t\, M_{t+1}R_{t+1}^e = 0$.*

9.5.3 The Equity Premium Puzzle*

Remark 9.17 *(Stein's lemma) If x and y have a bivariate normal distribution and $h(y)$ is a differentiable function such that $\mathrm{E}[|h'(y)|] < \infty$, then $\mathrm{Cov}[x, h(y)] = \mathrm{Cov}(x, y)\,\mathrm{E}[h'(y)]$.*

With CRRA utility, $c^{1-\gamma}/(1-\gamma)$, the SDF is

$$M_2 = \delta(c_2/c_1)^{-\gamma}$$

If the excess return, R^e, and consumption growth, Δc, have a bivariate normal distribution, then by using Stein's lemma, we can rewrite the risk premium (9.23) as (again dropping time subscripts)

$$\mathrm{E}\, R^e = \mathrm{Cov}(R^e, \Delta c)\gamma \tag{9.24}$$

$$= \mathrm{Corr}(R^e, \Delta c)\,\mathrm{Std}(R^e)\,\mathrm{Std}(\Delta c)\gamma. \tag{9.25}$$

The "equity premium puzzle" is that, over a long U.S. sample of the equity market and consumption per capita, $\mathrm{E}\, R^e \approx 0.08$, $\mathrm{Corr}(R^e, \Delta c) \approx 0.15$, $\mathrm{Std}(R^e) \approx 0.2$ and $\mathrm{Std}(\Delta c) \approx 0.02$, so an implausibly high risk aversion ($\gamma \approx 133$) is required to account for the high risk premia on the equity market. Basically, consumption is not volatile enough to explain the risk premium. See Cochrane (2005) for an extensive analysis.

Proof of (9.24). Stein's lemma gives $\mathrm{Cov}[R^e, \exp(\ln M)] = \mathrm{Cov}(R^e, \ln M) \, \mathrm{E}\, M$. (In terms of Stein's lemma, $x = R^e$, $y = \ln M$ and $h() = \exp()$.) Finally, notice that $\mathrm{Cov}(R^e, \ln M) = -\gamma \, \mathrm{Cov}(R^e, \Delta c)$. □

9.5.4 From a Consumption-Based Model to CAPM

Suppose the marginal utility (or stochastic discount factor) in equilibrium, is an affine function of the market excess return

$$U'(c) = a - bR^e_m, \text{ with } b > 0. \tag{9.26}$$

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. From (9.23) some rearrangements we get

$$\mathrm{E}\, R^e_i = \beta_i \, \mathrm{E}\, R^e_m, \text{ where } \beta_i = \sigma_{im}/\sigma^2_m, \tag{9.27}$$

which is the standard CAPM expression. This means that CAPM is consistent with (some) multi-period utility based portfolio choice models.

Proof of (9.27). Using (9.26) in (9.19) gives $\mathrm{E}\, R^e_i = b\sigma_{im} / \mathrm{E}(a - bR^e_m)$. We can, of course, apply this expression to the market excess return (instead of asset i) to get $\mathrm{E}\, R^e_m = b\sigma^2_m / \mathrm{E}(a - bR^e_m)$. Solve for $b / \mathrm{E}(a - bR^e_m)$ and use that in the first equation to get (9.27). □

9.5.5 From a Consumption-Based Model to a Multi-Factor Model

The consumption-based model is really a one-factor model, in terms of consumption. However, the relevant consumption level can be tricky to measure: both the treatment of durables and the identification of those who are actively investing are non-trivial. However, it may be reasonable to assume that we can approximate this (hard to observe) factor with a linear combination of other variables. Many macro models would suggest that a small number of (more easily measured) variables can provide an approximation. Also, the models may have missing factors that could be correlated with the state of the economy. For instance, marginal utility (or the stochastic discount factor) in equilibrium could be

$$U'(c) = ay + br, \tag{9.28}$$

where y denotes output and r the real interest rate.

It is then possible to write (9.23) as

$$\mathrm{E}\, R_i^e = \beta_{iy}\mu_y^e + \beta_{ir}\mu_r^e, \tag{9.29}$$

where (β_{iy}, β_{ir}) are from a multiple regression of R_{it}^e on excess returns on assets that are perfectly correlated with y and r respectively ("factor mimicking portfolios"), while (μ_y^e, μ_r^e) are the corresponding average excess returns. (The proof is in the Appendix.) The more general insight is that when the marginal utility (or stochastic discount factor) is linear in K factors, then we get a K-beta model for average returns.

9.6 Testing Multi-Factors Models

Let R_{ot}^e be a vector of factor *excess returns*. Testing whether $\alpha = 0$ in the regression

$$R_{it}^e = \alpha + \beta' R_{ot}^e + \varepsilon_{it} \tag{9.30}$$

is then the key approach to assess the model. (This test is invalid if some factors are not excess returns.)

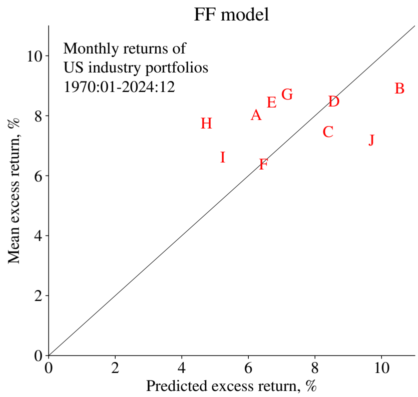
The t-test of the null hypothesis that $\alpha_i = 0$ uses the fact that, under fairly mild conditions, the t-statistic has an asymptotically normal distribution, that is

$$\frac{\hat{\alpha}_i}{\mathrm{Std}(\hat{\alpha}_i)} \overset{d}{\to} N(0,1), \tag{9.31}$$

under the null hypothesis ($\alpha_i = 0$). The standard error could be based on iid assumption or (better) account for heteroskedasticity.

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). Although this three-factor model is rejected at traditional significance levels, it still captures a fair amount of the variation of expected returns and may thus be a useful model.

Remark 9.18 *(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 risk-free return), since they are long-short portfolios. He and Ng (1994) find that SMB is related to*



	α (ann.)	t-stat	σ (ann.)
A (NoDur)	1.75	1.51	8.48
B (Durbl)	-1.68	-0.73	16.78
C (Manuf)	-0.98	-1.17	6.11
D (Enrgy)	-0.14	-0.07	15.93
E (HiTec)	1.70	1.30	9.52
F (Telcm)	-0.12	-0.08	10.73
G (Shops)	1.51	1.23	8.93
H (Hlth)	2.96	2.07	10.41
I (Utils)	1.34	0.88	11.04
J (Other)	-2.58	-3.36	5.60

Fama-French model
Factors: US market, SMB (size), and HML (book-to-market)
Predicted excess return: $\beta_m \bar{R}^e_m + \beta_{SMB} \bar{R}_{SMB} + \beta_{HML} \bar{R}_{HML}$

10% crit. value (Bonferroni): 2.58

Test if all $\alpha_i = 0$:

Wald stat	23.28
5% crit val	18.31
p-value	0.01

Figure 9.6: Fama-French regressions on US industry indices

macroeconomic risks, but HML less so.

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. For such (non-return) factors it is common to use factor mimicking portfolios: the excess return on portfolios strongly correlated with the factors.

Empirical Example 9.19 *Figure 9.6 shows some results for the Fama-French model on US industry portfolios and Figures 9.7–9.8 on the 25 Fama-French portfolios, both for more recent samples than in the original articles. The results indicate that the FF model is a considerable improvement compared to CAPM for the 25 FF portfolios, but perhaps not so much for the industry portfolios. Even for the 25 FF portfolios, strict statistical tests reject also the FF model, but the fit of the average returns is clearly better than CAPM.*

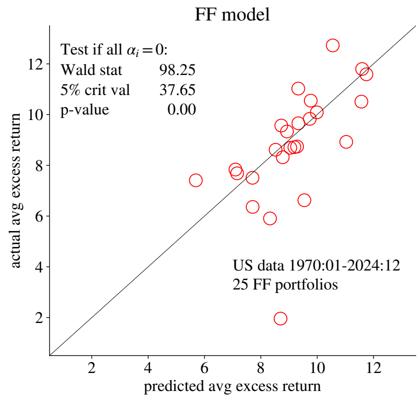


Figure 9.7: Fama-French regressions on FF portfolios

9.7 Appendix – The Asset Pricing Implications*

Proof (of (9.9)). Assume k in (9.8) is such that w equals the (financial) "market" portfolio, w_m in (9.8). For any portfolio with portfolio weights w_p, the covariance with the market return (times $1-\phi$) is

$$\begin{aligned}(1-\phi)\sigma_{pm} &= (1-\phi)w_p'\Sigma w_m \\ &= w_p'\Sigma\Sigma^{-1}\left(\mu^e/k - S_c\phi\right) \text{ using (9.8)} \\ &= \mu_p^e/k - \phi\sigma_{pc},\end{aligned}$$

since $w_p'\mu^e = \mu_p^e$ and $w_p'S_c = \sigma_{pc}$. Rearrange to put μ_p^e/k on the LHS and rewrite as

$$\mu_p^e/k = \begin{bmatrix}1-\phi & \phi\end{bmatrix}\begin{bmatrix}\sigma_{pm} \\ \sigma_{pc}\end{bmatrix} \tag{*}$$

Recall that $\sigma_{p\lambda} = \sigma_{pc}$ where λ indicates the factor mimicking portfolio with excess return R_λ^e (see Remark 9.9) and rewrite as

$$\begin{aligned}\mu_p^e/k &= \begin{bmatrix}1-\phi & \phi\end{bmatrix}\begin{bmatrix}\sigma_{mm} & \sigma_{m\lambda} \\ \sigma_{m\lambda} & \sigma_{\lambda\lambda}\end{bmatrix}\begin{bmatrix}\sigma_{mm} & \sigma_{m\lambda} \\ \sigma_{m\lambda} & \sigma_{\lambda\lambda}\end{bmatrix}^{-1}\begin{bmatrix}\sigma_{pm} \\ \sigma_{p\lambda}\end{bmatrix} \\ &= \begin{bmatrix}1-\phi & \phi\end{bmatrix}\begin{bmatrix}\sigma_{mm} & \sigma_{m\lambda} \\ \sigma_{m\lambda} & \sigma_{\lambda\lambda}\end{bmatrix}\begin{bmatrix}\beta_{pm} \\ \beta_{p\lambda}\end{bmatrix},\end{aligned} \tag{**}$$

*

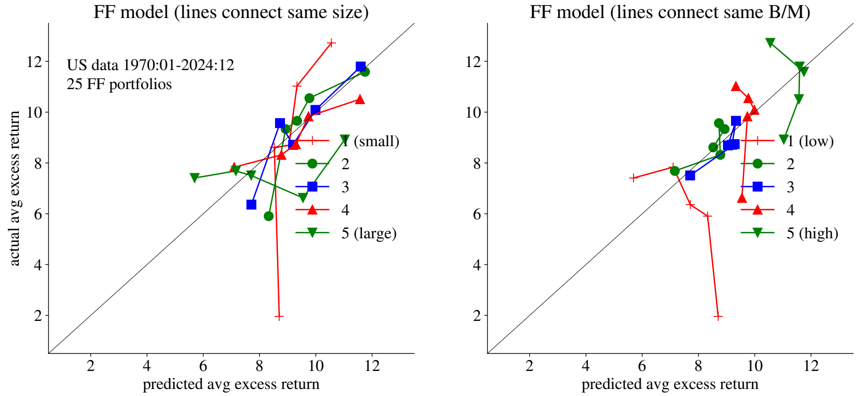


Figure 9.8: Fama-French regressions on FF portfolios

where $(\beta_{pm}, \beta_{p\lambda})$ are the coefficients from regressing R_p on (R^e_m, R^e_λ). For the market return when $p = m$, $(\beta_{mm}, \beta_{m\lambda}) = (1, 0)$, so (**) gives

$$\mu^e_m / k = \begin{bmatrix} 1-\phi & \phi \end{bmatrix} \begin{bmatrix} \sigma_{mm} \\ \sigma_{m\lambda} \end{bmatrix}.$$

For the factor mimicking portfolio when $p = \lambda$, $(\beta_{\lambda m}, \beta_{\lambda\lambda}) = (0, 1)$, so (**) gives

$$\mu^e_\lambda / k = \begin{bmatrix} 1-\phi & \phi \end{bmatrix} \begin{bmatrix} \sigma_{m\lambda} \\ \sigma_{\lambda\lambda} \end{bmatrix}.$$

Use these last two equations to substitute for the first two terms on the RHS of (**) and cancel the $1/k$ factors to get

$$\mu^e_p = \begin{bmatrix} \mu^e_m & \mu^e_\lambda \end{bmatrix} \begin{bmatrix} \beta_{pm} \\ \beta_{p\lambda} \end{bmatrix},$$

which is (9.9). $\square$

Proof (of (9.29)). Rewrite

$$\mathrm{E}\, R^e_i = \frac{a\sigma_{iy} + b\sigma_{ir}}{-\mathrm{E}(ay + br)} = \frac{1}{-\mathrm{E}(ay + br)} \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \sigma_{iy} \\ \sigma_{ir} \end{bmatrix}$$

Recall that $\sigma_{i\upsilon} = \sigma_{iy}$ and $\sigma_{i\rho} = \sigma_{ir}$, where R^e_υ (R^e_ρ) is the factor mimicking portfolio of y (r), see Remark 9.9). Rewrite as

$$\mathrm{E}\, R_i^e = \frac{1}{-\mathrm{E}(ay+br)} \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \sigma_{\upsilon\upsilon} & \sigma_{\upsilon\rho} \\ \sigma_{\upsilon\rho} & \sigma_{\rho\rho} \end{bmatrix} \begin{bmatrix} \sigma_{\upsilon\upsilon} & \sigma_{\upsilon\rho} \\ \sigma_{\upsilon\rho} & \sigma_{\rho\rho} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{i\upsilon} \\ \sigma_{i\rho} \end{bmatrix}$$
$$= \frac{1}{-\mathrm{E}(ay+br)} \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \sigma_{\upsilon\upsilon} & \sigma_{\upsilon\rho} \\ \sigma_{\upsilon\rho} & \sigma_{\rho\rho} \end{bmatrix} \begin{bmatrix} \beta_{i\upsilon} \\ \beta_{i\rho} \end{bmatrix}. \qquad (+)$$

Apply (+) to R_υ^e to get $(\beta_{i\upsilon}, \beta_{i\rho}) = (1, 0)$, so

$$\mu_y^e = \frac{1}{-\mathrm{E}(ay+br)} \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \sigma_{\upsilon\upsilon} \\ \sigma_{\upsilon\rho} \end{bmatrix}.$$

Similarly, apply (+) to R_ρ^e to get $(\beta_{i\upsilon}, \beta_{i\rho}) = (0, 1)$, so

$$\mu_r^e = \frac{1}{-\mathrm{E}(ay+br)} \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \sigma_{\upsilon\rho} \\ \sigma_{\rho\rho} \end{bmatrix}.$$

Use these two expressions to rewrite (+) as

$$\mathrm{E}\, R_i^e = \begin{bmatrix} \mu_\upsilon^e & \mu_\rho^e \end{bmatrix} \begin{bmatrix} \beta_{i\upsilon} \\ \beta_{i\rho} \end{bmatrix}.$$

□

Chapter 10

Efficient Markets

This chapter gives an overview of empirical methods for studying return predictability—along with results for some key assets. This can be used as inputs to different trading strategies, but also help formulating assumptions (about the average returns) needed for MV analysis and portfolio choice.

10.1 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, such as security analysis and market timing is of limited use and incurs costs (management fees, trading costs). Instead, it might be more practical to apply a passive approach that satisfies individual requirements (diversification, hedging background risk, appropriate risk level, etc). The implications of the EMH are thus significant.

10.1.1 Different Versions of the Classical Efficient Market Hypothesis

A precise formulation of the EMH needs to specify three things.

First, what type of information is used in making the forecasts? Is it price and trading volume data (referred to as 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 unpredictable? Most modern financial theory would focus on *excess returns*, since they represent risk compensation.

Third, what is the link between predictability and expectations? If an excess return is almost unpredictable, then *rational* investors would have nearly constant expected risk premia, and portfolio weights are likely to be fairly stable over time. The opposite holds if excess returns are straightforward to predict.

Rejection of the EMH, based on statistical studies of ex post samples, 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.

This chapter will present methods and empirical results. The first sections deal with traditional in-sample methods, initially focusing on the return history of the same asset, but later broadening the scope to bring in other types of predictors (fundamental valuation ratios, lagged returns of other assets, etc.) Later sections will instead focus on out-of-sample methods (recursive regressions, trading strategies, etc), and some evidence on the performance of professional forecasters.

10.2 Autocorrelations and Autoregressions

Autocorrelations and autoregressions are tools for studying whether past and current returns can predict future returns (typically of the same asset).

10.2.1 Autocorrelation Coefficients

The autocovariances of the R_t process can be estimated as

$$\hat{\gamma}_s = \textstyle\sum_{t=1+s}^{T} \left(R_t - \bar{R}\right)\left(R_{t-s} - \bar{R}\right)/T, \text{ where} \tag{10.1}$$

$\bar{R} = \Sigma_{t=1}^{T} R_t/T$ is the sample average estimated from the full sample. (In time series analysis we typically divide by T in (10.1) even if there are only $T-s$ observations to estimate γ_s from.) In most of the applications of this chapter, R_t indicates either a return or an excess return.

Autocorrelations are then estimated as

$$\hat{\rho}_s = \hat{\gamma}_s/\hat{\gamma}_0. \tag{10.2}$$

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 (for $s > 1$) are all zero, then for any lag s different from zero

$$\sqrt{T}\hat{\rho}_s \to^d N(0,1), \tag{10.3}$$

so $\sqrt{T}\hat{\rho}_s$ can be used as a t-stat.

Example 10.1 *(t-test) Reject the null hypothesis that $\rho_1 = 0$ on the 10% significance level if $\sqrt{T}|\hat{\rho}_1| > 1.64$.*

Empirical Example 10.2 *Figures 10.1–10.2 show autocorrelations for daily U.S. equity data.*

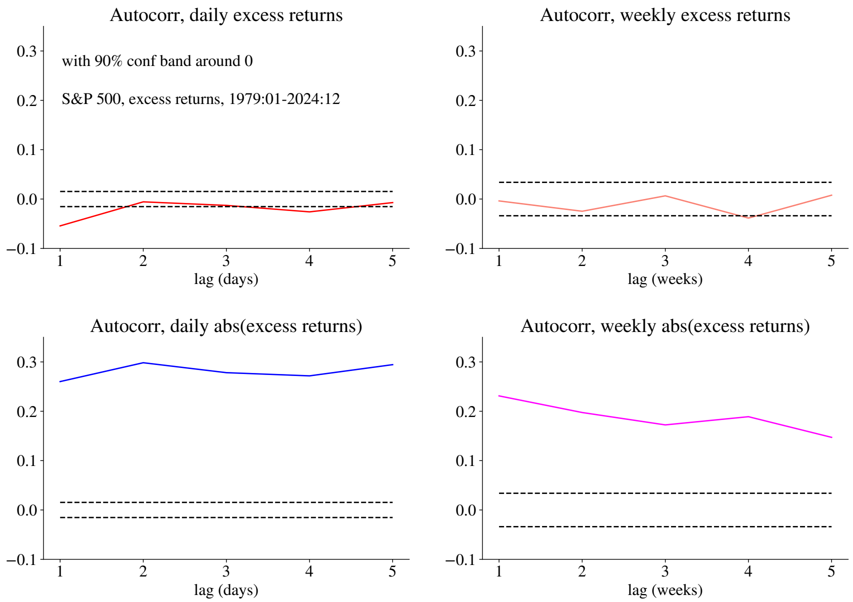


Figure 10.1: Predictability of US stock returns

10.2.2 Autoregressions

An alternative method for testing autocorrelations is to estimate an AR model

$$R_t = c + a_1 R_{t-1} + a_2 R_{t-2} + ... + a_p R_{t-p} + \varepsilon_t, \tag{10.4}$$

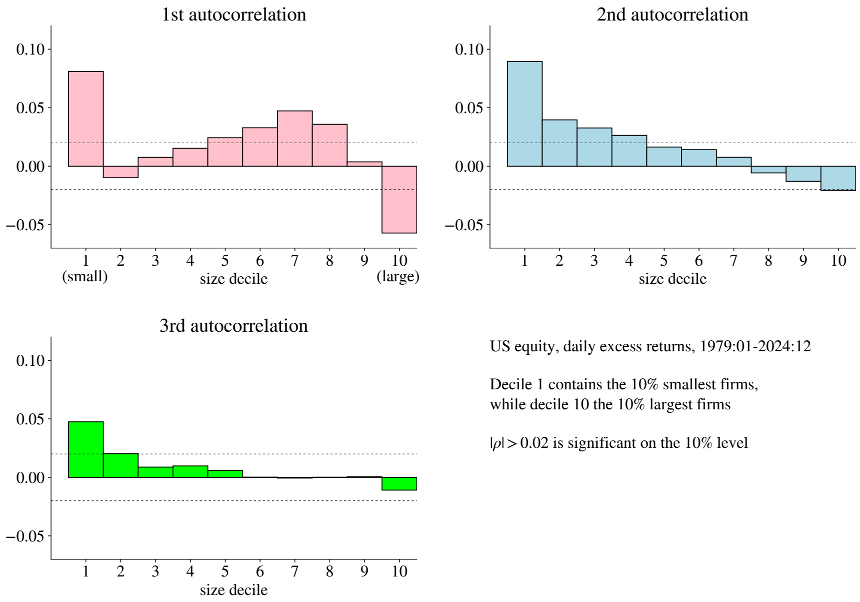


Figure 10.2: Predictability of US stock returns, size deciles

and 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.

Empirical Example 10.3 *Table 10.1 shows results from estimating an AR model on daily data for S&P 500 returns.*

The autoregression can also handle non-linear patterns. For instance, consider an AR(1), but where the autoregression coefficient may be different depending on the sign of last period's return

$$R_t = \alpha + \beta Q_{t-1} R_{t-1} + \gamma (1 - Q_{t-1}) R_{t-1} + \varepsilon_t, \text{ where} \tag{10.5}$$
$$Q_{t-1} = 1 \text{ if } R_{t-1} < 0 \text{ and } 0 \text{ otherwise.}$$

Empirical Example 10.4 *Figure 10.3 shows regression results from daily S&P 500 data. The reversal back after a negative shock is the most prominent finding.*

Empirical Example 10.5 *Figure 10.4 shows results from autoregressions for different investment horizons. For the business cycle frequency (3-4 years), there is some*

	Daily return
lag 1	−0.06
	(−2.73)
lag 2	−0.02
	(−0.70)
lag 3	−0.02
	(−0.95)
lag 4	−0.03
	(−1.95)
lag 5	−0.01
	(−0.53)
c	0.03
	(3.17)
R^2	0.01
All slopes	0.00
obs	10896

Table 10.1: AR(5) of daily S&P returns 1979:01-2024:12. 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.

evidence of negative autocorrelation, that is, reversals. However, testing long-run returns is challenging because it requires a very long sample to have enough (non-overlapping) return periods, and it is unclear if data from long ago is informative about today's economy.

The empirical evidence reported in this section suggest little autocorrelation for daily returns for large-cap stocks (like those in S&P 500), but perhaps more for smaller firms. There is also some indication of non-linearity, with more autocorrelation in down markets. For longer return horizons, there appear to be some negative autocorrelation on the business cycle frequency.

10.3 Other Predictors and Methods

There are many other possible predictors of future stock returns. For instance, lagged returns on other assets might predict returns, and both the dividend-price ratio and interest rates have been used to predict long-run returns.

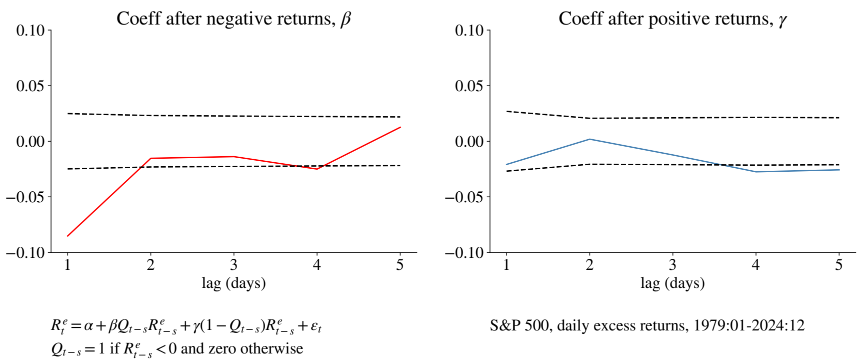


Figure 10.3: Predictability of US stock returns, results from a regression with interactive dummies

10.3.1 Lead-Lags

Stock indices have more positive autocorrelation than (most) individual stocks: there should therefore be fairly strong cross-autocorrelations among individual stocks. Indeed, this is also what is found in US data where returns of large size stocks forecast returns of small size stocks.

Empirical Example 10.6 *Figure 10.5 shows (for different size deciles) the regressions coefficients on the 1-day own lag and the 1-day lag of large caps. The results suggest considerable spillover from large caps to the other size deciles.*

10.3.2 Earnings-Price Ratio as a Predictor

One of the most effective methods of forecasting long-run returns is a regression of future returns on the current earnings-price (or dividend-price) ratio

$$R^e_{s,t} = \alpha + \beta_q \ln(\mathrm{e}_{t-s}/p_{t-s}) + \varepsilon_t, \tag{10.6}$$

where $R^e_{s,t}$ is the s-period excess return over the period $t-s$ to t.

Empirical Example 10.7 *Figure 10.6 shows results from estimating (10.6) for different investment horizons on data for a U.S. stock market index.*

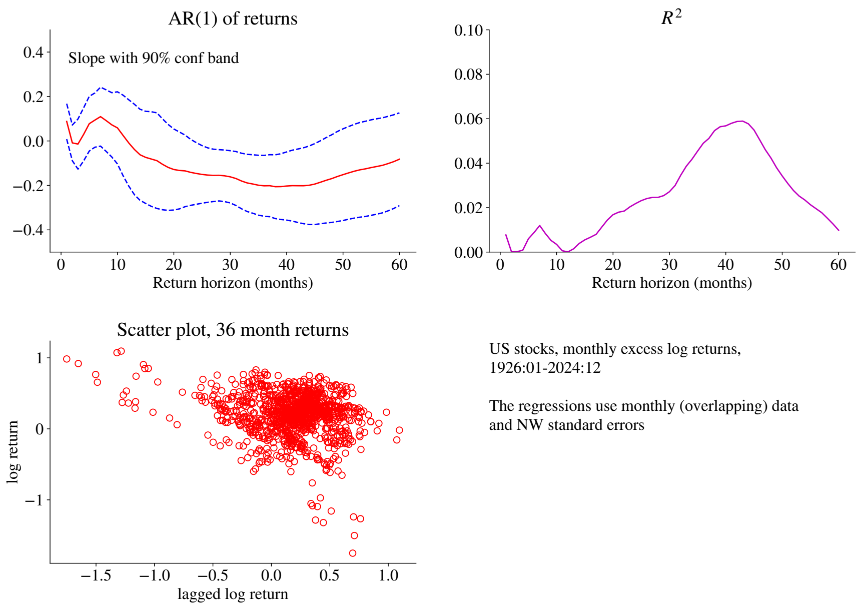


Figure 10.4: Predictability of long-run US stock returns

10.4 Out-of-Sample Forecasting Performance

10.4.1 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, forecasts are made using historical data and then updated as we get more information. To be precise, estimation is done using data up to and including $t-1$ and a forecast is made for t. Then we use data up to an including t to make a forecast for $t+1$ and so forth. This is called a recursive approach as the data sample is extended. See Figure 10.7. An alternative is to instead use a moving data window (ending in $t-1$ and then in t,...) where really old data points are discarded. Yet another approach is downweight old data. In either case, the forecasting performance is often compared with a benchmark model, for instance, using the historical average as the prediction estimated on the same sample.

One way to illustrate the relative forecast performance is the out-of-sample

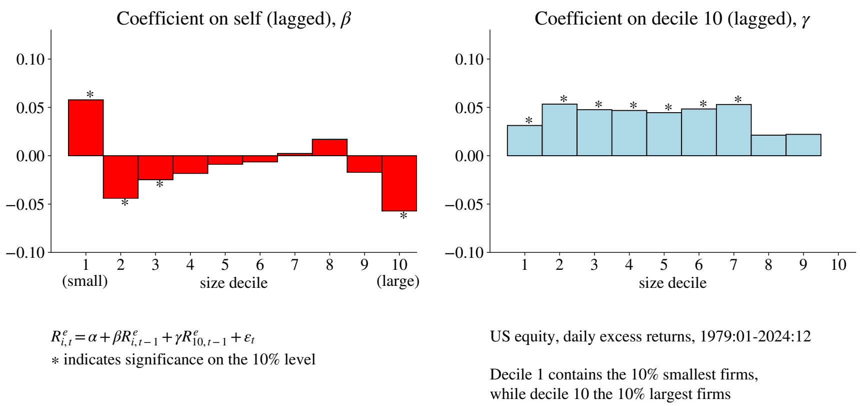


Figure 10.5: Coefficients from multiple prediction regressions

coefficient of determination (denoted R^2_{OS}, not to be confused with a return)

$$R^2_{OS} = 1 - \sum_{t=s}^{T} (R_t - \hat{R}_t)^2 / \sum_{t=s}^{T} (R_t - \tilde{R}_t)^2, \tag{10.7}$$

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 (typically also estimated on data up to and including $t-1$).

Example 10.8 *(R^2_{OS})*

$$R^2_{OS} = 1 - 0.4/0.5 = 0.2 \text{ (your model is better)}$$
$$R^2_{OS} = 1 - 0.5/0.4 = -0.25 \text{ (your model is worse)}$$

To statistically *test* the relative predictive performance, define

$$g_t = (R_t - \hat{R}_t)^2 - (R_t - \tilde{R}_t)^2, \tag{10.8}$$

and test if the sample average, $\bar{g}$, differs from zero, a method described by Diebold and Mariano (1995). Instead of squared forecast errors, we could consider absolute values or an indicator of whether the sign is right. If there is little or no autocorrelation in g_t, then $\text{Var}(\bar{g}) = \text{Var}(g_t)/T$ so the t-stat is

$$\frac{\bar{g}}{\text{Std}(g_t)/\sqrt{T}}, \tag{10.9}$$

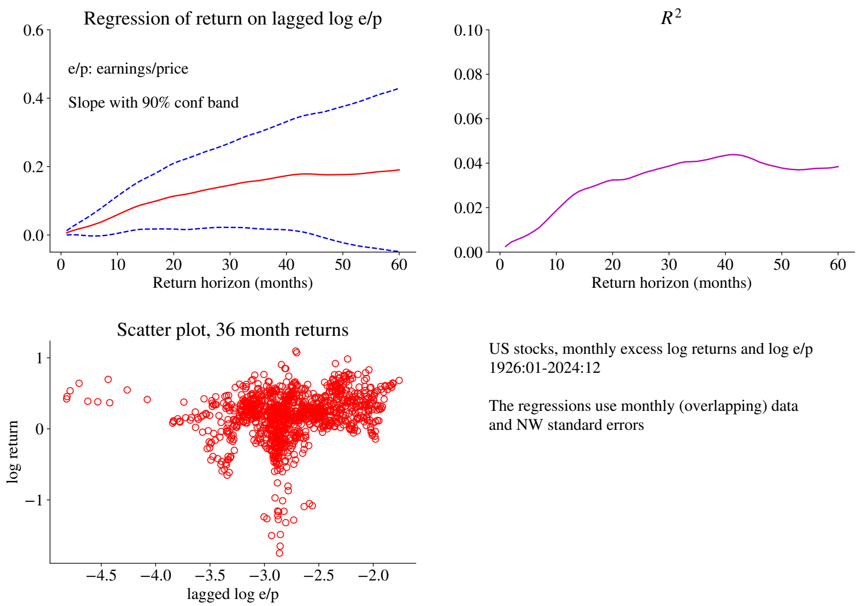


Figure 10.6: Predictability of long-run US stock returns

which could be compared with a $N(0, 1)$ distribution.

Empirical Example 10.9 *Figure 10.8 shows results based on daily data for different size deciles. It seems as if an AR(1) model is better than the historical average for small caps, but worse for large caps.*

Empirical Example 10.10 *Figure 10.9 shows how an e/p regression for a U.S. stock market index compares with the historical average—at different investment horizons. It seems as if it's consistently worse, which is similar to the findings of Goyal and Welch (2008).*

10.4.2 Trading Strategies

Another way to assess predictability and to illustrate its economic importance is to calculate the return of a *dynamic trading strategy*. In particular, the *alpha* (α) from a regression on the market excess return, $R^e_t = \alpha + \beta R^e_{mt} + \varepsilon_t$, is a useful measure. Neutral performance requires $\alpha = 0$, which can be tested with a t test.

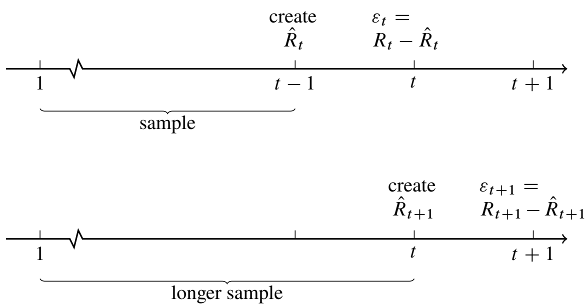


Figure 10.7: Out-of-sample forecasting

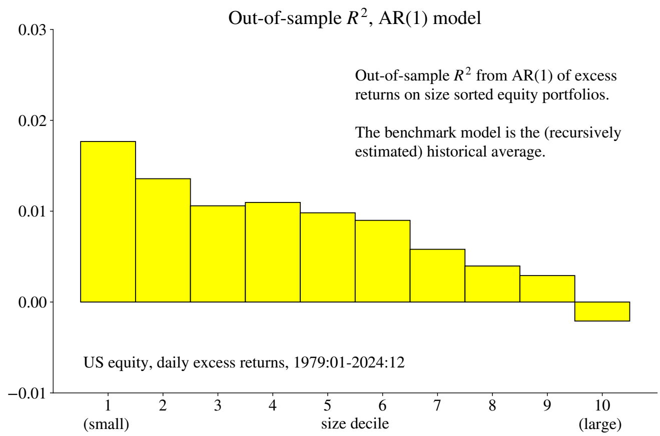


Figure 10.8: Short-run predictability of US stock returns, out-of-sample

Empirical Example 10.11 *See Figure 10.10 for an empirical example based on a momentum strategy (bet on recent winners, bet against recent losers) on daily data for the 25 FF portfolios. The upper left figure shows that the strategy has high average returns and α. It also shows that frequent rebalancing is important for the performance. The lower left figure illustrates the magnitude of trading costs that the strategy can handle, while still generating a positive average excess return. The upper right figure instead investigates the importance of a formation lag: having a time gap between the sorting and portfolio formation. The results suggest a short gap, perhaps shorter than a week.*

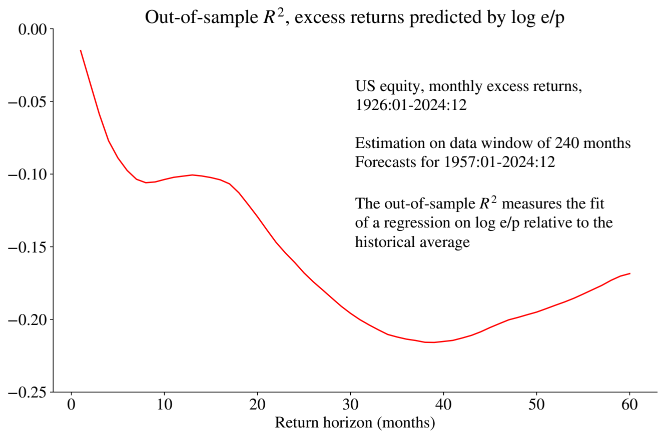


Figure 10.9: Predictability of long-run US stock returns, out-of-sample

10.4.3 Technical Analysis

Technical analysis is typically a data mining exercise which looks for local trends or systematic non-linear patterns (see, for instance, Brock, Lakonishok, and LeBaron (1992)). The basic idea is that markets are not instantaneously efficient, so prices may exhibit delayed and predictable reactions 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 models/trading rules.

Many trading rules rely on some kind of local trend which can be thought of as positive autocorrelation in price movements (also called momentum).

A *moving average rule* involves buying when a short moving average exceeds a long moving average. The idea is that this signals a new upward trend. Let S be the lag order of a short moving average and L of a long moving average, with $S < L$ and let b be a bandwidth (perhaps 0.01). Then, a MA rule for period t could be

$$\begin{bmatrix} \text{buy in } t \text{ if} & MA_{t-1}(S) > MA_{t-1}(L)(1+b) \\ \text{sell in } t \text{ if} & MA_{t-1}(S) < MA_{t-1}(L)(1-b) \\ \text{no change} & \text{otherwise} \end{bmatrix} \text{, where} \qquad (10.10)$$

$$MA_{t-1}(x) = (p_{t-1} + \ldots + p_{t-x})/x.$$

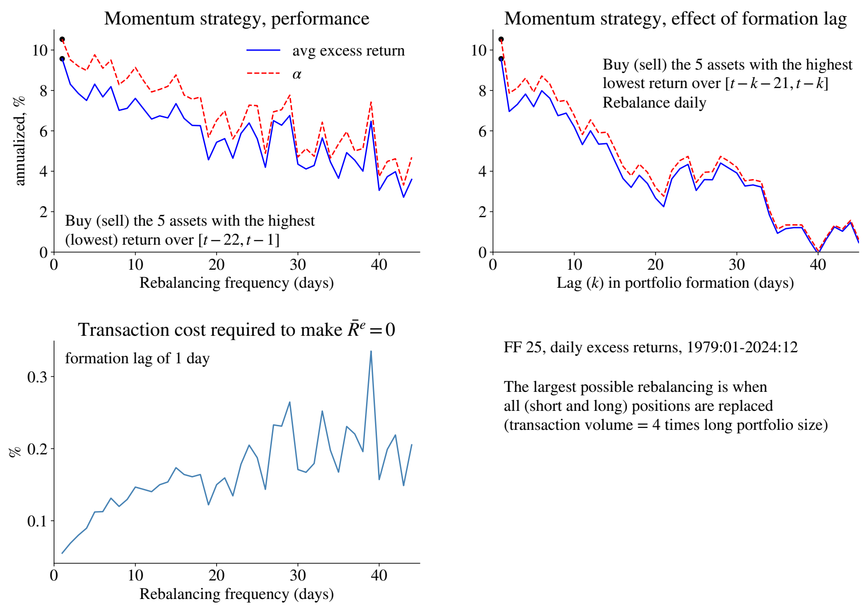


Figure 10.10: Predictability of US stock returns, momentum strategy

The difference between the two moving averages is called an *oscillator*

$$\text{oscillator}_t = MA_t(S) - MA_t(L), \tag{10.11}$$

(or sometimes, moving average convergence divergence, MACD) and the sign is often 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*[1], 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* generally involves 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 (round numbers often play the role as resistance levels). Once this (artificial?) resistance level has been broken, the price can possibly rise substantially.

[1]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).

On the downside, a *support level* plays the same role: some investors are willing to buy when the price reaches that value. To implement this, it is common to let the resistance/support levels be proxied by minimum and maximum values over a data window of length L. With a bandwidth b (perhaps 0.01), the rule for period t could be

$$\begin{bmatrix} \text{buy in } t \text{ if} & P_t > M_{t-1}(1+b) \\ \text{sell in } t \text{ if} & P_t < m_{t-1}(1-b) \\ \text{no change} & \text{otherwise} \end{bmatrix}, \text{ where} \tag{10.12}$$

$$M_{t-1} = \max(p_{t-1}, \ldots, p_{t-S})$$
$$m_{t-1} = \min(p_{t-1}, \ldots, p_{t-S}).$$

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 reverse the previous trading rules: we would typically sell when the price is high.

Empirical Example 10.12 *Figure 10.11 shows the idea of a reversal rule for S&P 500: buy when recent (a short MA) index values are outside the medium term trend (a long MA). The performance of implementing this rule over a long sample is shown in Table 10.2. The evidence suggest that the buy and sell signals do contain some information: average returns are high after buy signals, but that this may come at the cost of higher uncertainty (see Lo, Mamaysky, and Wang (2000) for a detailed study).*

10.5 Security Analysts

Reference: Makridakis, Wheelwright, and Hyndman (1998) 10 and Elton, Gruber, Brown, and Goetzmann (2014) 27

To do: update this section with more recent evidence

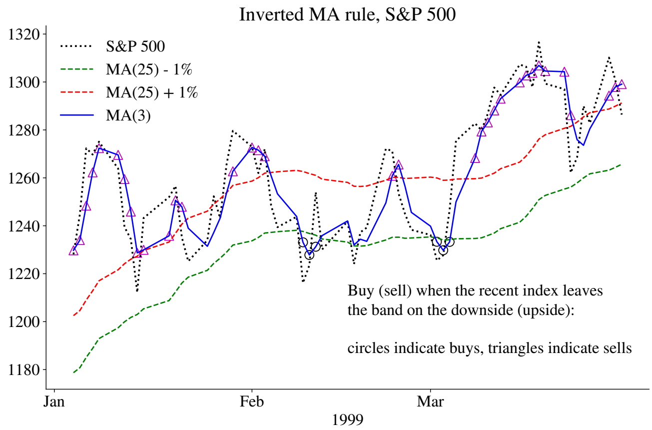


Figure 10.11: Example of a trading rule, illustration over short subsample

10.5.1 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.)

10.5.2 Do Security Analysts Overreact?

The paper by Bondt and Thaler (1990) compares the (semi-annual) forecasts (one- and two-year time horizons) with actual changes in earnings per share (1976-1984) for several hundred companies. The paper has regressions like

$$\text{Actual earnings change} = \alpha + \beta(\text{forecasted earnings change}) + \text{ residual},$$

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)

	Mean	Std
All days	7.2	18.0
After buy signal	16.2	27.3
After neutral signal	4.8	14.7
After sell signal	4.4	13.5
Strategy	9.1	27.5
Transaction cost	0.1	

Table 10.2: Excess returns (annualized, in %) from technical trading rule (Inverted MA rule). Daily S&P 500 data 1990:01-2024:12. The trading strategy involves (a) on every day: hold one unit of the index and short the risk-free; (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 the trade volume that the strategy can pay and still perform as well as the static holding of (a).

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.

10.5.3 High-Frequency Trading Based on Recommendations from Stock Analysts

Barber, Lehavy, McNichols, and Trueman (2001) give a somewhat different picture. They focus on the profitability of a trading strategy based on analyst recommendations. They use a huge data set (some 360,000 recommendations, US stocks) for the period 1985–1996. They sort stocks in to five portfolios depending on the consensus (average) recommendation—and redo the sorting every day (if a new recommendation is published). They find that such a daily trading strategy gives an annual 4% abnormal return on the portfolio of the most highly recommended stocks, and an annual -5% abnormal return on the least favourably recommended stocks.

This strategy requires a lot of trading (a turnover of 400% annually), so trading costs would typically reduce the abnormal return on the best portfolio to almost zero. A less frequent rebalancing (weekly, monthly) gives a very small abnormal return for the best stocks, but still a negative abnormal return for the worst stocks. Chance and Hemler (2001) obtain similar results when studying the investment advise by 30 professional "market timers."

10.5.4 Economic Experts

Several papers, for instance, Bondt (1991) and Söderlind (2010), have studied whether economic experts can predict the broad stock markets. The results suggests that they cannot. For instance, Söderlind (2010) shows that the economic experts that participate in the semi-annual Livingston survey (mostly bank economists) *(ii)* forecast the S&P worse than the historical average (recursively estimated), and that their forecasts are strongly correlated with recent market data (which in itself, cannot predict future returns).

10.5.5 Analysts and Industries

Boni and Womack (2006) study data on some 170,000 recommendations for a very large number of U.S. companies for the period 1996–2002. Focusing on revisions of recommendations, the papers shows that analysts are better at ranking firms within an industry than ranking industries.

10.5.6 Insiders

Corporate insiders *used to* earn superior returns, mostly driven by selling off stocks before negative returns. (There is little/no systematic evidence of insiders gaining by buying before high returns.) Actually, investors who followed the insider's registered transactions (in the U.S., these are made public six weeks after the reporting period), also used to earn some superior returns. It seems as if these patterns have more or less disappeared.

10.5.7 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.

Chapter 11

Performance Analysis

This chapter summarizes some performance measures commonly used to describe the past returns of mutual funds and other investment portfolios. The later sections also discusses performance attribution and style analysis.

11.1 Performance Evaluation

11.1.1 The Idea behind Performance Evaluation

Traditional performance analysis seeks to answer the question: "Should we include an asset in our portfolio, assuming future returns will follow the same distribution as in a historical sample."

Most performance measures rely on mean-variance (MV) analysis; however, the full MV portfolio optimization problem is not solved from scratch in these cases. 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. (A mix of the two is not considered.)

Although the analysis is based on the MV model, it does not assume that all portfolios conform to Capital Asset Pricing Model's (CAPM's) beta representation or that the market portfolio is the optimal choice for every investor. Rather, CAPM is used as an approximation.

Mutual fund evaluations typically find *(i)* neutral performance on average (or less due to trading costs and fees); *(ii)* poorer performance among large funds; *(iii)* better performance in less liquid and possibly less efficient markets; and *(iv)* little persistence in performance, except for very bad funds.

Example 11.1 *(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).*

Several popular performance measures are related to the CAPM regression

$$R_t^e = \alpha + \beta R_{mt}^e + \varepsilon_t, \tag{11.1}$$

where $\mathrm{E}\,\varepsilon_t = 0$ and $\mathrm{Cov}(R_{mt}^e, \varepsilon_t) = 0$. In many cases, R_{mt}^e represents the excess return on the market, but it could be some other benchmark return, for instance, for a segment of the market.

Example 11.2 *(Statistics for the examples of performance evaluations) The examples below use 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 11.1: Basic facts about the market and two other portfolios, α, β, and Std(ε) are from CAPM regression: $R_{it}^e = \alpha + \beta R_{mt}^e + \varepsilon_{it}$

11.1.2 Alpha

The intercept (α) in the regression (11.1) is often used as a performance measure. In CAPM, α measures the risk adjusted return. To see that, construct a portfolio with the weight β_i on the market portfolio (or some other benchmark) and $1 - \beta_i$ on the risk-free asset. The excess return on this portfolio is

$$R_p^e = \beta_i R_m^e, \tag{11.2}$$

since $R_p = \beta_i R_m + (1 - \beta_i) R_f$. This portfolio has the same systematic risk (sensitivity to the market) as asset i. As a practical matter, it is typically straightforward to create this portfolio by investing β_i in an index tracking fund and the remainder in the riskfee asset.

The α is then the difference in average excess returns of two portfolios with the same systematic risk

$$\alpha_i = \mathrm{E}\, R_i^e - \mathrm{E}\, R_p^e. \tag{11.3}$$

Empirical Example 11.3 *Table 11.2 shows the various performance measures for two large mutual funds. The Vanguard fund seems to perform best.*

	α	SR	$M_i^2 - M_m^2$	AR	Treynor	T^2
Market	0.00	0.40	0.00		7.19	0.00
Putnam	−0.28	0.37	−0.61	−0.07	6.83	−0.35
Vanguard	1.68	0.53	2.26	0.46	10.00	2.81

Table 11.2: Performance Measures of Putnam Asset Allocation: Growth A and Vanguard Wellington, weekly data 1999:01-2024:12 (annualized figures)

11.1.3 Sharpe Ratio and M^2

Suppose we want to determine whether fund p is better than fund q for allocating *all* our savings in. Again, we don't allow a mix of them. With MV preferences, the answer is that p is better if it has a higher Sharpe ratio—defined as

$$SR = \mu^e/\sigma. \tag{11.4}$$

Intuitively, for a given level of volatility, we obtain the highest expected return.

Example 11.4 *(Performance measure) From Example 11.2 we get the performance measures in Table 11.3.*

	SR	$M_i^2 - M_m^2$	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 11.3: Performance Measures

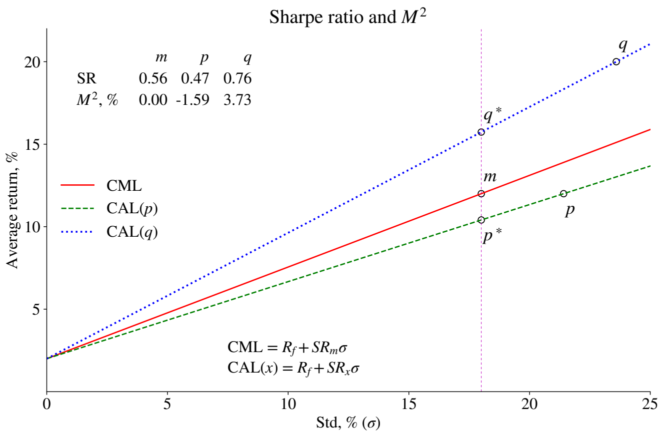


Figure 11.1: Sharpe ratio and M^2

Remark 11.5 *(Sortino ratio) The Sortino ratio is an alternative to the Sharpe ratio. It replaces σ_p with a measure of variation on the downside (typically, the square root of a semivariance).*

The M^2 (“Modigliani and Modigliani”) measure is

$$M^2 = R_f + SR_p\sigma_m, \tag{11.5}$$

which is simple transformation of the Sharpe ratio. The difference of M^2 for portfolio p and the market (or another benchmark) m can also be written as a difference of two risk-adjusted expected returns

$$\begin{aligned} M_p^2 - M_m^2 &= (SR_p - SR_m)\sigma_m \\ &= \mu_{p*}^e - \mu_m^e \end{aligned} \tag{11.6}$$

(or equivalently $\mu_{p*} - \mu_m$). In this expression, μ_{p*}^e is the expected excess return on a mix of portfolio p and the risk-free asset such that the volatility is the same as for the market return

$$R_{p*} = aR_p + (1-a)R_f, \text{ with } a = \sigma_m/\sigma_p. \tag{11.7}$$

The risk-adjustment here is thus to make the portfolios have the same volatility as the market. See Example 11.4 and Figure 11.1 for an illustration, which illustrate the relationship between the Sharpe ratio and M^2, highlighting the differences in risk-adjusted average returns.

Proof of (11.6). Notice that $SR_p = SR_{p^*}$ and that $\sigma_{p^*} = \sigma_m$. The first line of (11.6) can then be written $SR_{p^*}\sigma_{p^*} - SR_m\sigma_m$, which can be simplified as the 2nd line. □

11.1.4 Appraisal and Information Ratios

If the question 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 is

$$AR = \alpha / \operatorname{Std}(\varepsilon_t), \tag{11.8}$$

where α is the intercept and $\operatorname{Std}(\varepsilon_t)$ is the standard deviation of the residual ("tracking error") of the CAPM regression (11.1). If you think of (11.2) as the benchmark return, then AR is the average extra return per unit of extra standard deviation.

The motivation for AR indicating the best addition to the market portfolio is that the tangency portfolio based on *both* the market portfolio and portfolio p, has the following squared Sharpe ratio

$$SR_T^2 = AR^2 + SR_m^2. \tag{11.9}$$

(The proof is found below.) This is clearly increasing in AR (if positive), see Example 11.4 for an illustration.

The *information ratio*

$$IR_p = \frac{\mathrm{E}(R_p - R_b)}{\operatorname{Std}(R_p - R_b)}, \tag{11.10}$$

where R_b is some benchmark return. The information ratio is similar to both the Sharpe ratio and the appraisal ratio. The denominator in (11.10) can be thought of as the tracking error relative to the benchmark—and the numerator as the average active return (the gain from actively deviating from the benchmark). In fact, when the benchmark is as in (11.2), 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 of (11.9). From the CAPM regression (11.1) we have

$$\operatorname{Cov}\left(\begin{bmatrix} R_i^e \\ R_m^e \end{bmatrix}\right) = \begin{bmatrix} \beta_i^2\sigma_m^2 + \operatorname{Var}(\varepsilon_{it}) & \beta_i\sigma_m^2 \\ \beta_i\sigma_m^2 & \sigma_m^2 \end{bmatrix}, \text{ and } \begin{bmatrix} \mu_i^e \\ \mu_m^e \end{bmatrix} = \begin{bmatrix} \alpha_i + \beta_i\mu_m^e \\ \mu_m^e \end{bmatrix}.$$

As usual, the square of the Sharpe ratio of the tangency portfolio is $\mu^{e\prime}\Sigma^{-1}\mu^{e}$. Combining, we get that the squared Sharpe ratio for the tangency portfolio (using both R_{it} and R_{mt}) is

$$\left(\frac{\mu_T^e}{\sigma_T}\right)^2 = \frac{\alpha_i^2}{\mathrm{Var}(\varepsilon_{it})} + \left(\frac{\mu_m^e}{\sigma_m}\right)^2.$$

□

11.1.5 Treynor's Ratio and T^2

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

$$TR = \mu^e/\beta. \tag{11.11}$$

The basic intuition is that, with a *diversified portfolio* and a *small investment*, idiosyncratic risk becomes negligible, whereas only systematic risk (β) remains significant.

The TR measure can be rephrased in terms of expected returns—and could then perhaps be called the T^2 measure

$$\begin{aligned} T^2 &= \mu_p^e/\beta_p - \mu_m^e \\ &= \mu_{p*}^e - \mu_m^e \end{aligned} \tag{11.12}$$

(or equivalently, $\mu_{p*} - \mu_m$). In this expression, μ_{p*}^e is the expected excess return on a mix of portfolio p and the risk-free asset such that the beta is one (the same as for the market return)

$$R_{p*} = aR_p + (1-a)R_f, \text{ with } a = 1/\beta_p, \tag{11.13}$$

so $\mu_{p*}^e = \mu_p^e/\beta_p$. The risk-adjustment here is thus to make the portfolios have the same β as the market. See Example 11.4 and Figure 11.2 for an illustration which illustrate Treynor's Ratio and T^2, highlighting the differences in risk-adjusted average returns.

11.1.6 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 α. By using the expected values from

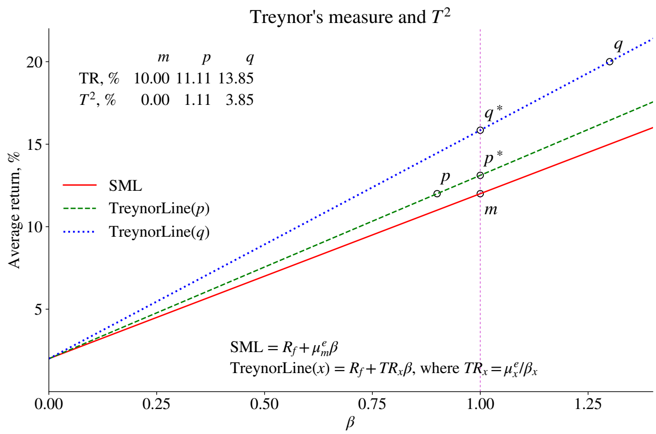


Figure 11.2: Treynor's ratio

the CAPM regression (11.2), $\mu_p^e = \alpha_p + \beta_p \mu_m^e$, simple rearrangements give

$$SR_p = \frac{\alpha_p}{\sigma_p} + \text{Corr}(R_p, R_m) SR_m$$
$$AR_p = \frac{\alpha_p}{\text{Std}(\varepsilon_{pt})}$$
$$TR_p = \frac{\alpha_p}{\beta_p} + \mu_m^e. \tag{11.14}$$

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 alpha is the driving force in all these measurements, this further motivates its use as a performance measure in itself.

Proof of (11.14). Taking expectations of the CAPM regression (11.1) gives $\mu_p^e = \alpha_p + \beta_p \mu_m^e$, where $\beta_p = \text{Cov}(R_p, R_m)/\sigma_m^2$. The Sharpe ratio is therefore

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

which can be written as in (11.14) since

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

The AR_p in (11.14) 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. □

11.1.7 More Sophisticated Performance Measures

This section goes beyond CAPM to consider more sophisticated performance measures.

The logic of using α from a CAPM regression can be extended to a *multi-factor model* where the factors are excess returns

$$R_t^e = \alpha + \beta_m R_{mt}^e + \beta_c R_{ct}^e + ... + \varepsilon_t. \tag{11.15}$$

Once again α can be seen as a performance measure and the rest (excluding ε_t) as a portfolio that could be easily replicated.

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 analogous to having a beta that varies linearly with the market excess return

$$\beta = b + cR_{mt}^e. \tag{11.16}$$

Using this in a traditional market model (CAPM) regression, $R_t^e = a_i + \beta R_{mt}^e + \varepsilon_t$, gives

$$R_t^e = a + bR_{mt}^e + c(R_{mt}^e)^2 + \varepsilon_t, \tag{11.17}$$

where c captures the ability to "time" the market. That is, if the investor systematically exits the market prior to periods of 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 as the benchmark, then $a + c(R_{mt}^e)^2$ could 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.

A recent way to merge the ideas of market timing and multi-factor models is to allow the coefficients to be time-varying according to some predetermined information ("state") variable, z_{t-1}. To illustrate this, suppose z_{t-1} is a single variable, so the time-varying (or "conditional") CAPM regression is

$$\begin{aligned} R_t^e &= (\theta_1 + \theta_2 z_{t-1}) + (\theta_3 + \theta_4 z_{t-1})\, R_{mt}^e + \varepsilon_t \\ &= \theta_1 + \theta_2 z_{t-1} + \theta_3 R_{mt}^e + \theta_4 z_{t-1} R_{mt}^e + \varepsilon_t. \end{aligned} \tag{11.18}$$

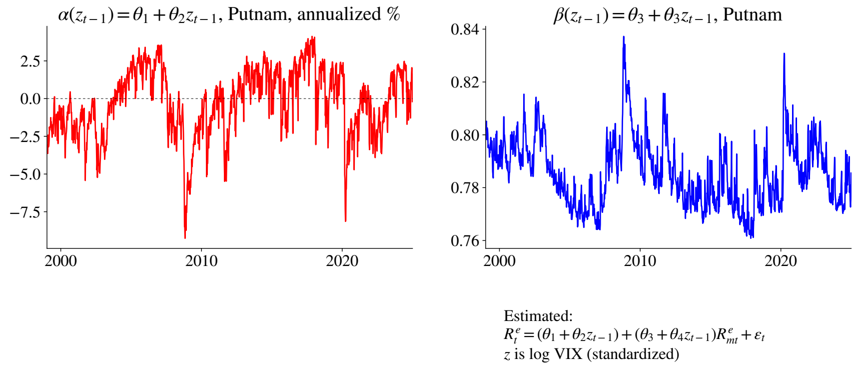


Figure 11.3: Conditional CAPM regression

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/risk-free asset), then only the first two terms are performance. In either case, the performance is time-varying.

Empirical Example 11.6 *(Conditional CAPM regression) Figure 11.3 illustrates the results from estimating (11.18) for the Putnam fund, using the logarithm of VIX (standardized to have zero mean and unit variance) as the state variable. The results indicate modest swings in the effective β, but considerable movements in the effective α.*

11.2 Holdings-Based Performance Measurement

As a complement to the purely return-based performance measurements discussed, it may be of interest to study how the portfolio weights change (if that information is available). This highlights *how* the performance has been achieved.

The Grinblatt-Titman measure (see Grinblatt and Titman (1993)) in period t is

$$GT_t = \sum_{i=1}^{n} (w_{i,t-1} - w_{i,t-2}) R_{it}, \tag{11.19}$$

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_{it} 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. Researchers commonly report the time-series average of GT_t.

11.3 Performance Attribution

The performance of an investment fund often depends on decisions taken on several levels of the organisation. To get a better understanding of where the performance was generated, a performance attribution calculation can be helpful. It uses information on portfolio weights (often in-house information) to decompose overall performance according to a number of criteria (typically related to different levels of decision making).

For instance, we could decompose the return 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

$$R_p = \textstyle\sum_{i=1}^{n} w_i R_{pi} \text{ and } R_b = \textstyle\sum_{i=1}^{n} v_i R_{bi}, \tag{11.20}$$

where w_i is the weight on asset class i (for instance, 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

$$R_p - R_b = \underbrace{\textstyle\sum_{i=1}^{n} (w_i - v_i) R_{bi}}_{\text{allocation effect}} + \underbrace{\textstyle\sum_{i=1}^{n} w_i \left(R_{pi} - R_{bi}\right)}_{\text{selection effect}}. \tag{11.21}$$

The first term is the *allocation effect*, that is, the importance of allocation across asset classes, measured using the benchmark return of that asset class. If decisions on allocation to different asset classes are taken by senior management (or a board), then this is the contribution of that level. Instead, the second term is the *selection effect*, that is, the importance of selecting the individual securities within an asset class, as it depends on difference in returns that the fund and the benchmark earns in a given asset class. This contribution is more likely to come from the the trading desk.

Remark 11.7 *(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 (11.21) since* $\sum_{i=1}^{n}(w_i - v_i)R_b = R_b\sum_{i=1}^{n}(w_i - v_i) = 0$.

11.3.1 What Drives Differences in Performance across Funds?

Much 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), see for instance, Ibbotson and Kaplan (2000). For other investors, including hedge funds, leverage also plays a role.

11.4 Style Analysis

Style analysis (Sharpe (1992)) is a way to use econometric tools to find out the portfolio composition from a series of the returns, at least in broad terms. This is clearly a bit cruder than having access to the actual portfolio weights (as dicussed above), especially since the estimation requires some data points before the portfolio change can be detected.

The key idea is to identify several return indices (typically 5 to 10) believed to account for the majority of the portfolio's returns, followed by running regressions to find the portfolio "weights." It is essentially a multi-factor regression without a intercept and where the coefficients are constrained to sum to unity (and, optionally, to be positive)

$$\begin{aligned} R_{pt}^{e} &= \textstyle\sum_{j=1}^{K} b_j R_{jt}^{e} + \varepsilon_{pt}, \text{ with} \\ \textstyle\sum_{j=1}^{K} b_j &= 1 \text{ and } b_j \geq 0 \text{ for all } j. \end{aligned} \tag{11.22}$$

Clearly, the restrictions could be changed to $U_j \leq b_j \leq L_j$, which could allow for some short positions.

The coefficients are typically estimated by minimizing the sum of squared residuals. In case the only restriction is that the coefficient should sum to one, then this can be solved with basic linear algebra (see the Remark below). With restrictions on the individual coefficient (for instance, no short sales), this is a non-linear least squares problem, but there are very efficient numerical methods for such problems.

Remark 11.8 *(Restricted OLS*) If we want to impose the restrictions $Rb = q$ on OLS where R is an $L \times K$ matrix and q is an $L \times 1$ vector, then the closed form solution is*

$$\hat{b} = b_{OLS} - S_{xx}^{-1} R' (R S_{xx}^{-1} R')^{-1} (R b_{OLS} - q),$$

where $S_{xx} = \Sigma_{t=1}^{T} x_t x_t'$ (that is, $X'X$ if row t of X contains x_t') and b_{OLS} is the unrestricted OLS estimate. For instance, $R = \mathbf{1}_K'$ and $q = 1$ gives the style analysis solution (11.22) except that short sales are allowed.

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.

Empirical Example 11.9 *See Figure 11.4 for an example of style analysis for the two mutual funds studied earlier. The results indicate that the coefficients move considerably over time (based on estimations from rolling data windows) and that the R^2 is above 95% except for the first few years.*

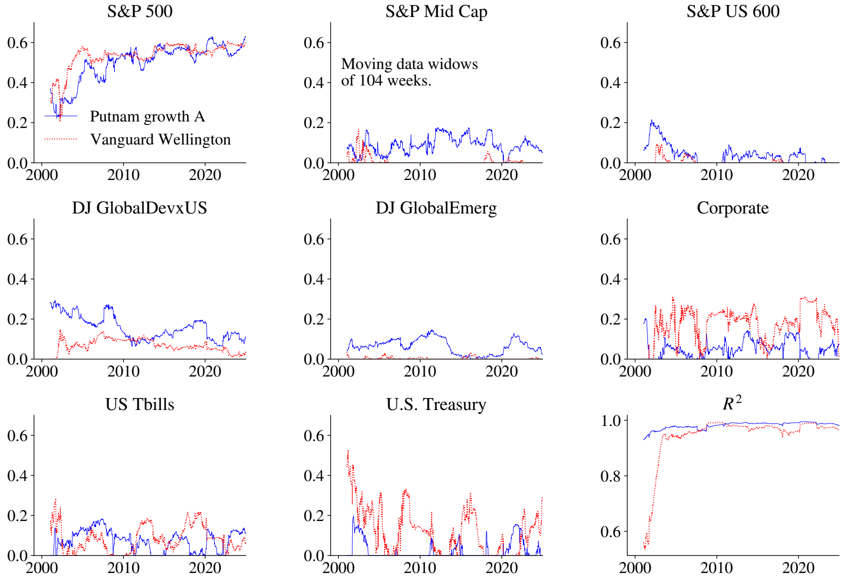


Figure 11.4: Example of style analysis, rolling data window

Chapter 12

Investment for the Long Run

This chapter studies long-run portfolio choices to find that the autocorrelation pattern of returns is of key importance. In particular, mean reversion of the asset price level (negative autocorrelation of the returns) turns out to be the most important factor that potentially could make equity safer for a long-run investor than for a short-run investor. Other arguments are weaker.

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: *(1)* Sharpe ratios increase with the investment horizon and *(2)* the probability that equity returns outperform bond returns increases with the horizon.

This chapter will compare these findings with results from mean-variance (MV) analysis.

Empirical Example 12.1 *Figure 12.1 shows how, for the U.S. equity market index, the Sharpe ratio and the probability of outperforming a safe asset differ across investment horizons.*

12.1.1 Long-Run Return as a Sum of Short-Run Returns

This section explains how a long-run return can be expressed in terms of multiple short-run returns. In particular, we use logarithmic returns, since they can easily be cumulated over time.

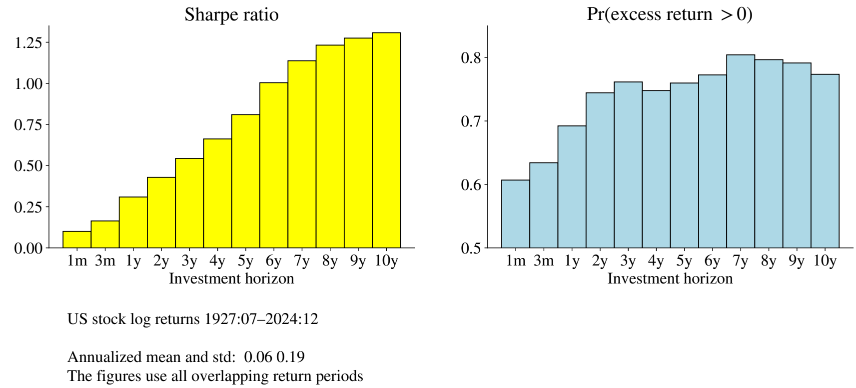


Figure 12.1: Empirical evidence on SR and probability of excess return > 0

The gross (buy-and-hold) return on a q-period investment made in period 0 can be written

$$1 + Z(q) = \prod_{t=1}^{q}(1 + R_t), \tag{12.1}$$

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

$$z(q) = \sum_{t=1}^{q} r_t, \tag{12.2}$$

where $z(q) = \ln(1 + Z(q))$ and where the log one-period return is $r_t = \ln(1 + R_t)$. 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 $z^e(q)$.

Remark 12.2 *(Approximating q-period returns*) It is sometimes convenient to approximate the q-period net return $Z(q)$ as*

$$Z(q) \approx \sum_{t=1}^{q} R_t.$$

This approximation works well unless there are numerous periods or extreme one-period returns. For instance, if $R_1 = 0.9$ and $R_2 = -0.9$ (indeed very extreme returns), then the two-period net return is $Z(2) = (1 + 0.9)(1 - 0.9) - 1 = -0.81$, while the approximation gives $Z(2) \approx R_1 + R_2 = 0$. The difference 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: this difference is much smaller.

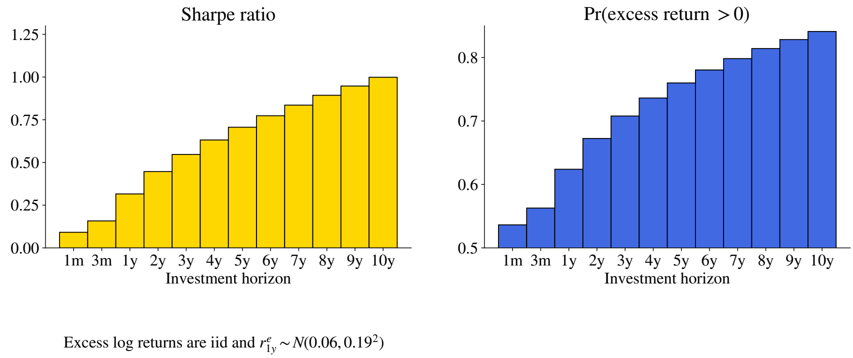


Figure 12.2: SR and probability of excess return > 0, iid returns

Remark 12.3 *(Geometric mean returns*) The average log return,* $\bar{r} = \Sigma_{t=1}^{q} r_t/q$*, is closely related to the geometric mean return. To see that, notice that a geometric mean gross return is* $1 + \tilde{R} = [\Pi_{t=1}^{q}(1 + R_t)]^{1/q}$*, which equals* $\exp(\bar{r})$*. For values close to 0,* $\tilde{R} \approx \bar{r}$*.*

12.1.2 Increasing Sharpe Ratios

This section demonstrates that with iid (independently and identically distributed) one-period returns, both the average and variance of a multi-period return grow linearly with the investment horizon. Consequently, the Sharpe ratio, defined as the expected excess return divided by the standard deviation, increases with the square root of horizon. This needs to be considered when comparing Sharpe ratios across investment horizons.

As before, let $z(q)$ be the log return on a q-period investment. *If log returns are iid*, the Sharpe ratio of $z(q)$ is

$$SR(z(q)) = \sqrt{q}SR(r), \tag{12.3}$$

where $SR(r)$ is the Sharpe ratio of the *one-period* log return. Clearly, this scales with the horizon, q. See Figure 12.2 for an illustration.

Proof of (12.3). 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 $\mathrm{E}\, z^e(q) = q\mu^e$ and $\mathrm{Var}(z(q)) = q\sigma^2$. $\square$

12.1.3 Probability of Outperforming a Risk-Free Asset

Under the assumption of normally distributed returns, the increasing Sharpe ratios imply higher probabilities of out-performing a risk-free asset.

In particular, assume that the log one-period returns are jointly normally distributed, which carries over to the q-period excess log return, $z^e(q)$. Then, we have

$$\Pr(z^e(q) > 0) = \Phi[SR(z(q))], \tag{12.4}$$

where $\Phi()$ is the cumulative distribution function of a standard normal variable, $N(0,1)$. Again, see Figure 12.2 for an illustration.

Together with the results in (12.3) this suggests that the empirical evidence in Figure 12.1 could potentially be explained by iid returns.

Proof of (12.4). By standard manipulations we have that if $x \sim N(\mu, \sigma^2)$, then

$$\Pr(x \leq 0) = \Pr((x - \mu)/\sigma \leq -\mu/\sigma) = \Phi(-\mu/\sigma),$$

since $(x - \mu)/\sigma$ is an $N(0,1)$ variable. Clearly, $\Pr(x > 0) = 1 - \Pr(x \leq 0)$. Use the fact that $\Phi(z) + \Phi(-z) = 1$ (since the standard normal distribution is symmetric around zero) and substitute $z^e(q)$ for x (and notice that μ/σ then corresponds to a Sharpe ratio) to get (12.4). $\square$

12.1.4 Why These Arguments Are Not Enough

Although increasing Sharpe ratios (at longer investment horizons) suggest a higher probability of out-performing a risk-free asset, that does not necessarily imply that the risky asset is safer for a long-run investor. We also have to take into account the size of the loss—in case the asset underperforms. With a longer horizon (and therefore higher dispersion), *really* bad outcomes are more likely: 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 further explore how the investment horizon affects the portfolio weights, it is necessary to clarify the investor's preferences, specifically how risks and opportunities are compared.

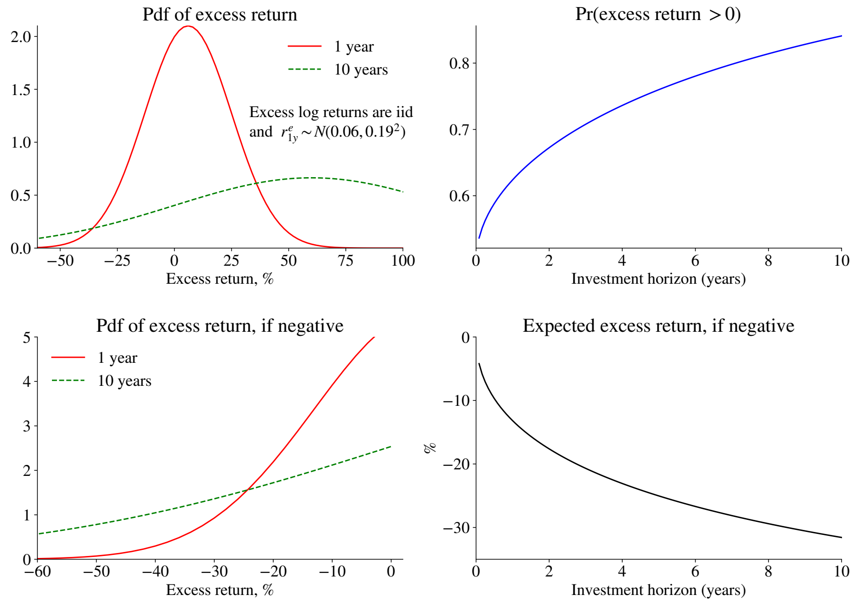


Figure 12.3: Time diversification, normally distributed returns

12.2 Mean-Variance Portfolio Choice

12.2.1 Approximating the Log Portfolio Return

Logarithmic portfolio returns are convenient in a dynamic setting since they are additive across time. However, they have a drawback on the portfolio formation stage: the logarithmic portfolio return is a *non-linear* function of the logarithmic returns of the assets. Therefore, we will use an approximation.

If there is only one risky asset with return Z and risk-free asset with return Z_f, then the portfolio return is $Z_p = vZ + (1-v)Z_f$. For this case we approximate the log portfolio return as

$$z_p = \ln[ve^z + (1-v)e^{z_f}] \tag{12.5}$$

$$\approx z_f + v(z - z_f) + v\sigma_z^2/2 - v^2\sigma_z^2/2, \tag{12.6}$$

where σ_z^2 is the variance of z over the relevant investment horizon (see Campbell and Viceira (2002)). As usual, all moments represent the beliefs of the investor, conditional on the information available at the time of investment. For convenience,

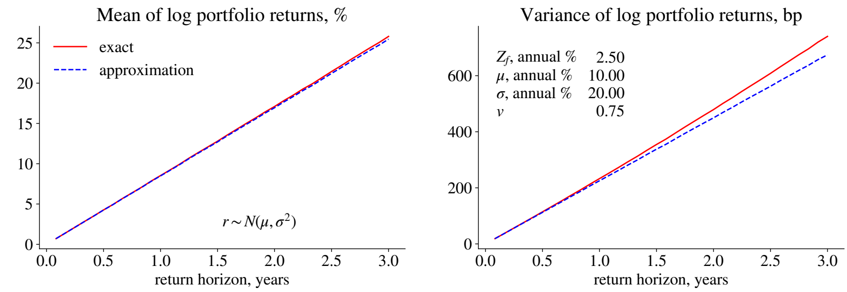


Figure 12.4: Mean and variance of log portfolio return for different return horizons, exact and according the approximation (12.6)

we suppress the indicator q for the investment horizon.

The approximation error from (12.5) is straightforward to calculate, and it will differ across return levels. What is more important, however, is how the mean and variance differ between the exact calculation and the approximation. To assess that we need to assume a distribution of the returns. Figure 12.4 provides an illustration, where the basic assumption is that the log return on the risky asset is normally distributed, with a mean and standard deviation that are broadly in line with the U.S. equity market index. The figure suggests that both the mean and variance scale fairly linearly with the return horizon, and that the approximation works well up to 2–3 years. After that, we see a deviation, where the approximation underestimates the variance somewhat so a manual adjustment of the risk aversion parameter k might be sensible.

Proof of (12.6). The portfolio return $Z_p = vZ + (1-v)Z_f$ can be used to write

$$(1+Z_p)/(1+Z_f) = 1 + v[(1+Z)/(1+Z_f) - 1].$$

The logarithm is

$$z_p - z_f = \ln\{1 + v[\exp(z - z_f) - 1]\}.$$

The function $f(x) = \ln\{1 + v[\exp(x) - 1]\}$, where $x = z - z_f$, has the following derivatives (evaluated at $x = 0$): $df(x)/dx = v$ and $d^2 f(x)/dx^2 = v(1-v)$, and notice that $f(0) = 0$. A second order Taylor approximation of the log portfolio return around $z - z_f = 0$ is then

$$z_p - z_f \approx v(z - z_f) + v(1-v)(z - z_f)^2/2.$$

In a continuous time model, the square would equal its expectation, σ_z^2, so this further approximation is used to give (12.6). □

12.2.2 Mean-Variance Optimization

We assume that the investor solves a traditional mean-variance problem, but expressed in terms of log returns

$$\max_v \operatorname{E} z_p(q) - \frac{k}{2} \operatorname{Var}(z_p(q)), \tag{12.7}$$

where $z_p(q)$ is the q-period log return of a portfolio of a risky and a risk-free asset. Notice that we further assume that the investor picks a portfolio and then stays with it, that is, we rule out rebalancing during the investment horizon. (A later chapter will relax that.)

Using the approximation (12.6), the optimal weight on the risky asset is

$$v = \frac{\mu_z^e(q)}{(1+k)\sigma_z^2(q)} + \frac{1}{2(1+k)}. \tag{12.8}$$

This is, of course, very similar to the traditional MV results based on net returns. In particular, the key driver is the mean excess return divided by the variance and the risk aversion.

Example 12.5 *(of (12.8)) With* $(\mu_z^e, \sigma_z^2, k) = (0.008, 0.05^2, 5)$ *we get* $v \approx 0.62$*, but with* $(\mu_z^e, \sigma_z^2) = (0.016, 0.0594^2)$ *we get* $v \approx 0.84$.

Proof of (12.8). Using (12.6), (12.7) is approximately the same as

$$\max_v z_f + v\mu_z^e + v\sigma_z^2/2 - v^2\sigma_z^2/2 - kv^2\sigma_z^2/2,$$

where z_f is the risk-free rate over the investment horizon and (μ_z^e, σ_z^2) are mean and variance of the excess log return of the risky asset. (The indicator q for the investment horizon is suppressed.) The first order condition is $\mu_z^e + \sigma_z^2/2 - v(1+k)\sigma_z^2 = 0$, which gives (12.8). □

12.2.3 Mean-Variance Optimization with iid Logarithmic Returns

When log returns are iid, then both the mean and the variance scale with the investment horizon

$$\mu_z^e(q) = q \operatorname{E} r^e \text{ and } \sigma_z^2(q) = q \operatorname{Var}(r), \tag{12.9}$$

where $\operatorname{E} r^e$ is the expected excess log 1-period return and $\operatorname{Var}(r)$ its variance.

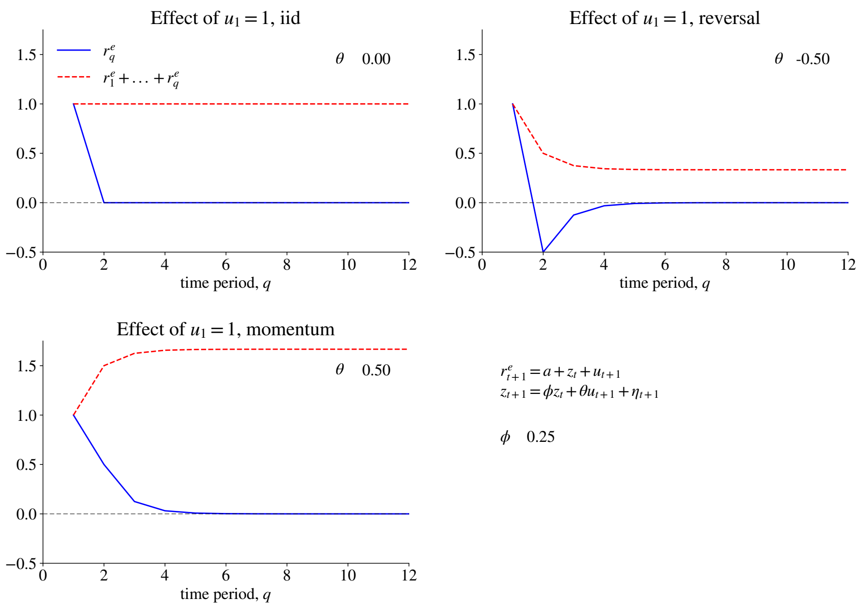


Figure 12.5: Impulse responses to an u_1 shock in the time series model (12.10)

This implies that, with iid log returns, the portfolio weight on the risky asset (12.8) is the same for all investment horizons, q. This is in stark contrast to what the increasing Sharpe ratios and probability of beating a risk-free asset would suggest. The key reason is that the mean-variance preferences consider also the magnitude of a loss, not just the probability of one.

12.2.4 A Time Series Model for Autocorrelated Logarithmic Returns

To discuss the case of autocorrelated returns, we use a simple time series model for 1-period log returns which allows for predictability

$$\begin{bmatrix} r^e_{t+1} \\ z_{t+1} \end{bmatrix} = \begin{bmatrix} a \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & \phi \end{bmatrix} \begin{bmatrix} r^e_t \\ z_t \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \theta & 1 \end{bmatrix} \begin{bmatrix} u_{t+1} \\ \eta_{t+1} \end{bmatrix} \tag{12.10}$$

where u_{t+1} and η_{t+1} are iid and uncorrelated shocks, with standard deviations σ_u and σ_η, respectively. The z_t variable can be thought of as a state variable which affects future average returns. The θ parameter controls how return shocks u_{t+1} spill over to future average returns, which will turn out to be a key feature. Formally, (12.10) is a state space model with correlated shocks, but it is written on VAR(1) form. For our purposes, the model has the advantage that is can generate both long-run momentum and reversals. (In contrast, an AR(1) with a negative coefficient has an oscillating forecast for future values.)

With $\theta = 0$ and $\sigma_\eta = 0$ (no spillover from u to z and no shocks to z), the return process is iid, which is the case discussed above. In contrast, with $\theta < 0$ a positive return shock in $t + 1$ will tend to be followed by negative returns, that is, a long-run reversal. See Figure 12.5 for an illustration. The reversal lasts for one period in case there is no autoregression in z_t ($\phi = 0$), but over a sequence of periods when there is positive autocorrelation in z_t ($\phi > 0$). The figure illustrates the latter case. In particular, notice how the response of long-run returns (the sum of 1-period returns $r^e_1 + \ldots r^e_q$), is muted by the reversal. This will imply that the volatility of long-run returns is low. In contrast, $\theta > 0$ will cause movements in the same direction (momentum) and increase thus volatility, again see Figure 12.5.

Since the shocks are iid and uncorrelated, the expectation and variance of r^e_{t+1}, *conditional on the information available* at the time of investment in t, are straightforward to calculate (see the Appendix). Figure 12.6, which is roughly calibrated to monthly U.S. equity data although the autocorrelations are exaggerated to make a point, provides an illustration. In particular, it shows how the variance of q-period returns, conditional on the information available at the time of investment ($t = 0$ in the figure), scales with the investment horizon q when returns are iid, but slower when the model exhibits mean reversion ($\theta < 0$) and vice versa.

12.2.5 Mean-Variance Optimization with Autocorrelated Logarithmic Returns

Autocorrelation can affect both the expectations and the uncertainty of future returns. However, the analysis here will focus on how the uncertainty depends on the investment horizon, disregarding the "market timing" issue. That is, we assume a neutral *initial* state, $z_t = 0$, but allow for future shocks to the state.

In general, positive autocorrelation will make the sum of returns, $z(q)$, have a variance that scales quicker than the return horizon q as shocks "build up" over time. The opposite holds for a negative autocorrelation.

Figure 12.6 suggests that (12.10) can replicate the iid case (12.9). But it also shows that with long-run reversal, uncertainty increases slower than the investment horizon, so equity is safer for a long-run investor. See Figure 12.7 for an illustration

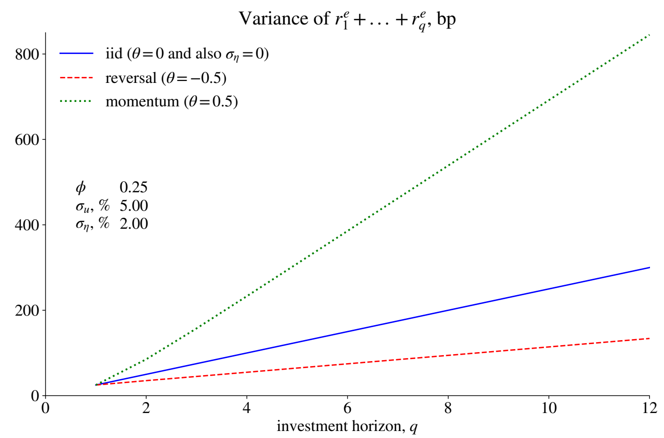


Figure 12.6: Variances of q-period return in the time series model (12.10)

of the optimal portfolio weight on the risky asset, based on the same rough calibration as before as well as the same exaggeration of the autocorrelation. The key result is that the optimal weight on the risky asset increases with the investment horizon—if the returns show reversals (are negatively autocorrelated).

In summary, this analysis suggests that iid log returns are *not* sufficient to make equity relatively safer for a long-run investor—if we use a mean-variance approach to portfolio choice. Rather, it requires long-run reversals. Empirical evidence suggests that there might be some reversals, but not very much, questioning the notion of equity being safe in the long run.

Note, however, that the analysis in this chapter relies on the assumption that the investor makes *one* portfolio choice, irrespective of investment horizon. That is, no rebalancing. A later chapter will look at that issue in more detail as well as discuss the optimal response to differences in the initial state.

12.2.6 Utility Based Portfolio Choice

To study whether the conclusions from the MV approach are robust, this section considers utility based portfolio choice. For instance, a *logarithmic utility function* means setting $k = 0$ in (12.7)–(12.8).

For a CRRA utility function and normally distributed log portfolio returns, we know (from an earlier chapter) that maximizing $\mathrm{E}(1 + Z_p)^{1-\gamma}/(1-\gamma)$ is equivalent

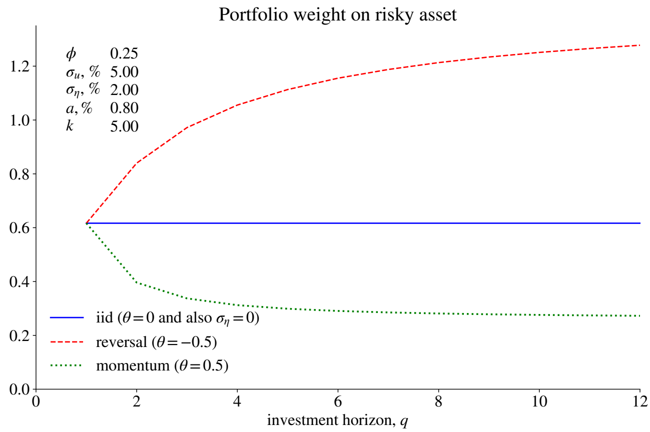


Figure 12.7: Portfolio weight on risky asset based on the time series model (12.10)

to maximizing

$$\mathrm{E}\, z_p - (\gamma - 1)\,\mathrm{Var}(z_p)/2, \tag{12.11}$$

which is once again of the same form as (12.7)–(12.8), but with $k = \gamma - 1$. (See the chapter on utility theory for a proof.)

Both these examples lend support to the conclusion from the MV approach: to make equity safer for a long-run investor, returns must show reversals, so the asset price level is mean-reverting.

12.3 Appendix – The Conditional Variances*

First, simulate the system by setting $u_t = 1$ and all other shock to zero. Trace out the dynamic effects on $r_t, r_{t+1}, ...$ Let the first element in ψ_0 be the effect on r_t, the first element in ψ_1 be the effect on r_{t+1}. Now, instead set $\eta_t = 1$ and trace of the dynamic effects to fill the second elements in the ψ vectors. We thus have the MA representation of a return in period t

$$r_t = \psi_0' \varepsilon_t + \psi_1' \varepsilon_{t-1} + \psi_2' \varepsilon_{t-2} + \ldots, \text{ where } \varepsilon_t = \begin{bmatrix} u_t \\ \eta_t \end{bmatrix}.$$

Second, notice that the innovations of future returns, compared to the information available in t, can be written

$$r_{t+1} = \begin{bmatrix} \mathbf{0}'\varepsilon_{t+3} \\ \mathbf{0}'\varepsilon_{t+2} \\ \psi_0'\varepsilon_{t+1} \end{bmatrix}, r_{t+2} = \begin{bmatrix} \mathbf{0}'\varepsilon_{t+3} \\ \psi_0'\varepsilon_{t+2} \\ \psi_1'\varepsilon_{t+1} \end{bmatrix}, r_{t+2} = \begin{bmatrix} \psi_0'\varepsilon_{t+3} \\ \psi_1'\varepsilon_{t+2} \\ \psi_2'\varepsilon_{t+1} \end{bmatrix}.$$

Note that shocks in t or earlier are known in t, so they do not enter the expressions. Also, future shocks cannot affect current returns, which explains the zeros.

Third, since ε_{t+1} is uncorrelated across time, variances and covariances are straightforward to calculate, for instance,

$$\mathrm{Cov}_t(r_{t+2}, r_{t+3}) = \psi_0'\Omega\psi_1 + \psi_1'\Omega\psi_2, \text{ where } \Omega = \mathrm{Var}(\varepsilon_t).$$

Fourth, once we have the variance-covariance matrix of $(r_{t+1}, r_{t+2}, r_{t+3})$, the variance of $r_{t+1} + r_{t+2} + r_{t+3}$ can be calculated as the sum of all the elements.

Chapter 13

Dynamic Portfolio Choice

This chapter discusses portfolio choice of a long-run investor who can rebalance in each period. This means that the investor may form a portfolio that also hedges (predictable) future changes of the investment opportunity set.

This will unfortunately give a more complicated optimization problem, so this chapter is somewhat harder reading than most of the preceeding chapters.

13.1 Logarithmic Utility

Let the objective in period t be to maximize the expected log wealth in some future period

$$\max \mathrm{E}_t \ln W_{t+q} = \max(\ln W_t + \mathrm{E}_t\, r_{p,t+1} + \mathrm{E}_t\, r_{p,t+2} + \ldots + \mathrm{E}_t\, r_{p,t+q}), \quad (13.1)$$

where r_{pt} is the log portfolio return, $r_{pt} = \ln(1 + R_{pt})$ with R_{pt} being the net portfolio return. The investor can rebalance the portfolio weights every period.

Remark 13.1 *(The Kelly criterion*) The portfolio that solves (13.1) is said to be "growth optimal" as it maximizes the expected growth of the portfolio, also known as the Kelly criterion. It can be noted that this portfolio also maximizes the geometric mean return. To see this, recall from an earlier chapter that the geometric mean return is an increasing function of the average log return. Maximizing one of them means maximizing the other.*

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 maximizes $\mathrm{E}_t\, r_{p,t+1}$.That is, to choose the one-period growth-optimal portfolio. This *myopic* approach is thus the optimal *dynamic* portfolio choice. Notice that the investment horizon q does not matter: short-run and long-run investors choose the same portfolio. This is specific to the logarithmic utility function.

However, the portfolio choice may change over time (t), if the distribution of the returns changes; that is, when returns are *not iid*, but this is unrelated to the investment horizon.

To solve the optimization problem, we approximate the *log* portfolio return, $r_p = \ln(1 + R_p)$, as in Campbell and Viceira (2002). (An earlier chapter includes a proof and an application.)

Remark 13.2 *(Approximate log portfolio return) The log portfolio return,* $\ln(1+)R_p = \ln(1 + v'R + (1 - v'\mathbf{1})R_f)$*, is approximately*

$$r_{pt} \approx r_{ft} + v'(r_t - r_{ft}) + v' \operatorname{diag}(\Sigma)/2 - v'\Sigma v/2, \tag{13.2}$$

where Σ is the $n \times n$ variance-covariance matrix of r_t and $\operatorname{diag}(\Sigma)$ *is the n-vector of the variances (that is, the diagonal elements of Σ). With a single risky asset, this can be simplified as*

$$r_{pt} \approx r_{ft} + v(r_t - r_{ft}) + v\sigma^2/2 - v^2\sigma^2/2, \tag{13.3}$$

where σ^2 is variance of r_t.

Using the approximation (13.2) and maximizing $\mathrm{E}_t\, r_{p,t+1}$ gives the optimal n-vector of portfolio weights as

$$v = \Sigma^{-1}(\mu^e + \operatorname{diag}(\Sigma)/2), \tag{13.4}$$

where μ^e is the vector of excess log returns of the risky assets, Σ is their variance-covariance matrix and $\operatorname{diag}(\Sigma)$ picks out the diagonal of Σ, that is, the vector of variances. (The proof is at the end of the section.) The weight on the risk-free asset is the remainder, $1 - v'\mathbf{1}$. The case of a single risky asset was solved in an earlier chapter, yielding $v = \mu^e/\sigma^2 + 1/2$.

Clearly, the portfolio weights v change over time if the expected excess returns and/or the variance-covariance matrix change; that is, when returns are not iid. We could think of this as a *managed portfolio*.

Proof of (13.4). From (13.2) we have that the objective function can be written $r_f + v'\mu^e + v' \operatorname{diag}(\Sigma)/2 - v'\Sigma v/2$, so the first order conditions are $\mu^e + \operatorname{diag}(\Sigma)/2 - \Sigma v = \mathbf{0}_{n\times 1}$, which gives (13.4). $\square$

Example 13.3 *(One risky asset) Suppose there is one risky asset with* $\sigma = 5\%$*, and the expected excess returns are different the two "scenarios" A and B:* $\mu^e_A = 0.8\%$ *or* $\mu^e_B = 0.2\%$*. Then (13.4) gives* $v_A = 3.7$ *and* $v_B = 1.3$ *in the two scenarios.*

Example 13.4 *(Three risky assets) Suppose we have three assets with the variance-covariance matrix (which is the same in both states)*

$$\Sigma = \begin{bmatrix} 83 & 17 & 29 \\ 17 & 32 & 2 \\ 29 & 2 & 50 \end{bmatrix} bp,$$

and the means (in scenario A and B, respectively)

$$\mu^e_A = \begin{bmatrix} 0.8 \\ 0.9 \\ 0.3 \end{bmatrix} \% \text{ and } \mu^e_B = \begin{bmatrix} 0.4 \\ 0.45 \\ 0.15 \end{bmatrix} \%,$$

In this case, the portfolio weights in the two states are

$$v_A \approx \begin{bmatrix} 0.65 \\ 2.93 \\ 0.60 \end{bmatrix} \text{ and } v_B \approx \begin{bmatrix} 0.49 \\ 1.62 \\ 0.45 \end{bmatrix}.$$

Empirical Example 13.5 *Figure 13.1 illustrates mean returns and standard deviations, estimated by exponentially weighted moving averages. Figure 13.2 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.*

13.2 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.

As a benchmark, recall that an earlier chapter has established that if the log portfolio return, $r_p = \ln(1 + R_p)$, is normally distributed, then maximizing E(1 +

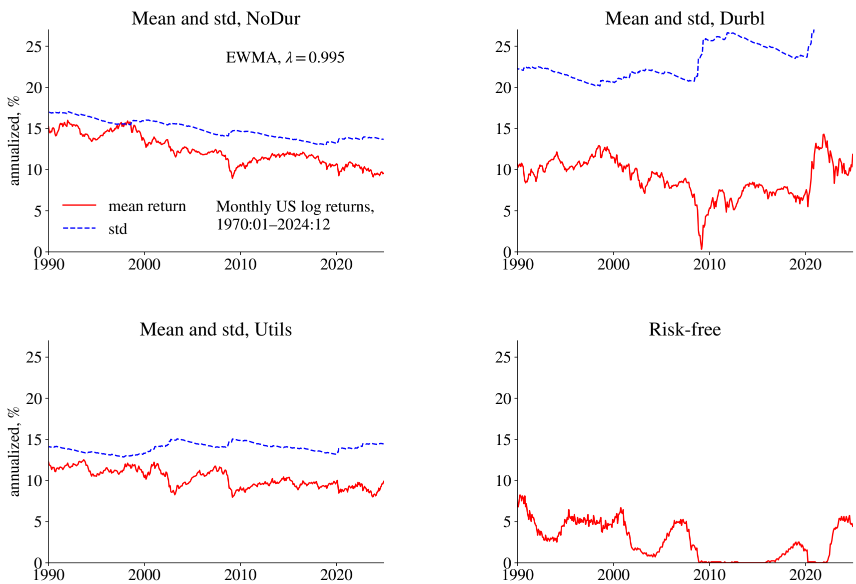


Figure 13.1: Dynamically updated estimates, 5 U.S. industries

$R_p)^{1-\gamma}/(1-\gamma)$ is equivalent to maximizing

$$\mathrm{E}\, r_p + (1-\gamma)\,\mathrm{Var}(r_p)/2. \tag{13.5}$$

Note that this is a one-period (myopic) optimum, since the utility function only depends on the return in the next periods. A dynamic optimum is discussed later on.

Using the approximation (13.2) gives optimal portfolio weights for the one-period (myopic) case as

$$v = \Sigma^{-1}(\mu^e + \mathrm{diag}(\Sigma)/2)/\gamma. \tag{13.6}$$

These are the same weights as from the log utility case, but now divided by the risk aversion γ.

Proof of (13.6). From (13.2) we have that the objective function can be written $r_f + v'\mu^e + v'\,\mathrm{diag}(\Sigma)/2 - v'\Sigma v/2 + (1-\gamma)v'\Sigma v/2$, so the first order conditions are $\mu^e + \mathrm{diag}(\Sigma)/2 - \gamma\Sigma v = \mathbf{0}_{n\times 1}$, which gives (13.6). $\square$

Example 13.6 *(One risky asset) Using the same figures as in Example 13.3 and* $\gamma = 6$ *gives* $v_A = 0.62$ *and* $v_B = 0.22$.

Example 13.7 *(Three risky assets) Using the same figures as in Example 13.4 and*

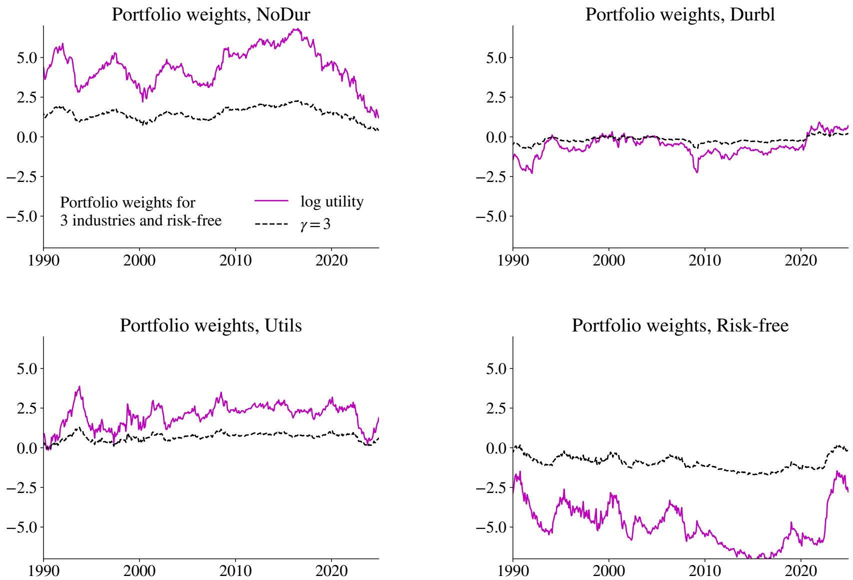


Figure 13.2: Dynamically updated portfolio weights, T-bill and 5 U.S. industries

$\gamma = 6$ *gives*

$$v_A \approx \begin{bmatrix} 0.11 \\ 0.49 \\ 0.10 \end{bmatrix} \text{ and } v_B \approx \begin{bmatrix} 0.08 \\ 0.27 \\ 0.07 \end{bmatrix}.$$

Empirical Example 13.8 *See Figure 13.2 for a comparison of the solutions from log utility and from CRRA with $\gamma = 3$. The changes are more muted with the higher risk aversion.*

13.3 Intertemporal Hedging

The *combination* of a CRRA utility function (with $\gamma \neq 1$) and non-iid returns makes the portfolio choice more challenging. If there is a link between returns in different periods, then a long-run investor might want to take this into account as it provides "diversification" across periods. This is *intertemporal hedging*. For instance, if some asset return (in $t + 1$) is negatively correlated with the investment outlook between $t + 1$ and $t + 2$, then that asset is (in t) seen as providing a hedge.

We illustrate this below by using a simple model, although there are a few steps involved in solving it. (See Merton (1973) and Campbell and Viceira (1999) for more elaborate approaches.) We also compare (in several figures) with myopic and static (multi-period) investment to highlight the differences.

13.3.1 An Autocorrelated Return Process

We use a simple time series model which encompasses both the iid case as well as long-run reversal or momentum

$$\begin{bmatrix} r^e_{t+1} \\ z_{t+1} \end{bmatrix} = \begin{bmatrix} a \\ \mathbf{0}_n \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{n\times n} & I_n \\ \mathbf{0}_{n\times n} & \phi \end{bmatrix} \begin{bmatrix} r^e_t \\ z_t \end{bmatrix} + \begin{bmatrix} I_n & \mathbf{0}_{n\times n} \\ \theta & I_n \end{bmatrix} \begin{bmatrix} u_{t+1} \\ \eta_{t+1} \end{bmatrix}. \tag{13.7}$$

where r^e and z (and thus also u and η) are n-vectors. We assume that u_{t+1} is uncorrelated with η_{t+1}, but there may be correlations within each vector. The respective variance-covariance matrices are Σ_{uu} and $\Sigma_{\eta\eta}$. Note that a is an n-vector and that ϕ and θ are both $n \times n$ matrices. (An earlier chapter used a similar model for the case of $n = 1$.)

13.3.2 Myopic Portfolio Choice

A *myopic investor* maximizes (13.5), which gives the same solution for the portfolio weights as in (13.6), but with

$$\mu^e = a + z_t \text{ and } \Sigma = \Sigma_{uu}. \tag{13.8}$$

Notice that both the expectation and variance are conditional on the information available at the time of the portfolio choice (t).

Figures 13.3–13.4 indicate the myopic portfolio weights with dots, mostly to make a comparison with the other cases discussed in more detail below.

Proof of (13.8). Notice that $r^e_{t+1} = a + z_t + u_{t+1}$. This immediately gives $\mathrm{E}_t\, r^e_{t+1} = a + z_t$ and $\mathrm{Var}_{t+1}(r^e_{t+1}) = \Sigma_{uu}$. $\square$

13.3.3 A Two-Period Investor (No Rebalancing)

We now consider a two-period investor who does not rebalance. This investor also maximizes (13.5), but the expectation and variance are for a 2-period return, so the

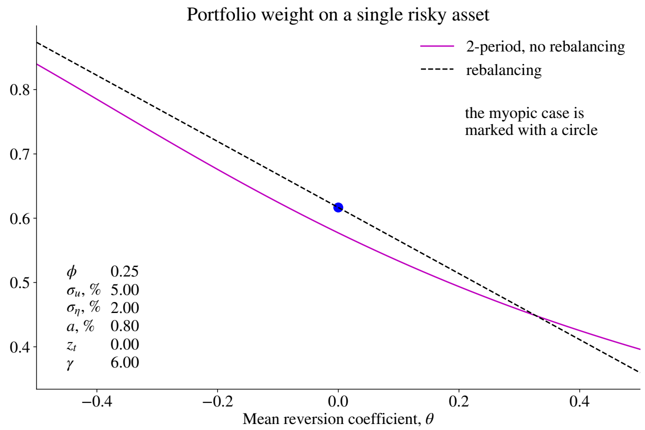


Figure 13.3: Weight on a single risky asset, two-period investor with CRRA utility for three cases: myopic, 2-period optimization without rebalancing and also with rebalancing. The return process is a scalar version of (13.7).

$\mathrm{E}\, r_p$ and $\mathrm{Var}(r_p)$ terms should (respectively) be interpreted as

$$\mathrm{E}_t\, r^e_{p,t+1} + \mathrm{E}_t\, r^e_{p,t+2} \text{ and} \tag{13.9}$$

$$\mathrm{Var}_t(r_{p,t+1}) + \mathrm{Var}_t(r_{p,t+2}) + 2\,\mathrm{Cov}_t(r_{p,t+1}, r_{p,t+2}). \tag{13.10}$$

(The second line defines $\mathrm{Var}_t(r_{p,t+1} + r_{p,t+2})$.) This involves, among other things, the covariance of the returns in the two periods, which is different from the myopic case. Intuitively, assets with reversals (negative autocorrelation) are less risky.

Again, we get the same solution for the portfolio weights as in (13.6), but with

$$\mu^e = 2a + (I + \phi)z_t \tag{13.11}$$

$$\Sigma = 2\Sigma_{uu}(I + \theta') + \theta\Sigma_{uu}\theta' + \Sigma_{\eta\eta}. \tag{13.12}$$

(See below for a proof.) When returns are iid ($\theta = \mathbf{0}, \Sigma_{\eta\eta} = \mathbf{0}$), then these expressions are two times those for the myopic case (13.8).

See Figure 13.3 for an illustration of how the portfolio weight on a single risky asset depends on the degree of reversal (θ, on the horizontal axis). In particular, with strong reversal ($\theta < 0$), equity is relatively safe for a long-run investor, which increases the portfolio weight. This is driven by the $\Sigma_{uu}\theta'$ term in (13.12), which captures the covariance of the returns in different periods. With reversal, this

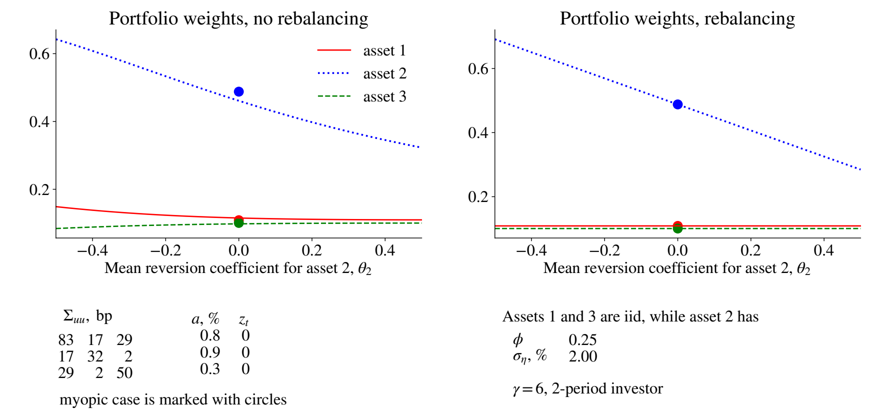


Figure 13.4: Weights on three risky asset, two-period investor with CRRA utility for three cases: myopic, 2-period optimization without rebalancing and also with rebalancing. The return process is a (13.7).

covariance is negative. (At $\theta = 0$ in the figure, the 2-period investor holds less risky assets than the myopic investor just because the volatility of z_{t+1}, $\Sigma_{\eta\eta} > 0$, adds uncertainty.) Notice that this figure assumes a neutral state ($z_t = 0$) in order to focus on the effect of risk.

Also, see Figure 13.4 for an illustration of a case with several risky assets (also assuming $z_z = 0$). It shows that the portfolio weight on the asset with non-iid returns (asset 2 in the figure) reacts strongly to variation in the reversal/momentum (here θ_2). Again, reversal makes the asset safer. There are some spillover effects on the other assets because of correlations.

Proof of (13.11)–(13.12). By combining the equations in (13.7) we can also write $r^e_{t+1} = a + z_t + u_{t+1}$ and $r^e_{t+2} = a + \phi z_t + \theta u_{t+1} + \eta_{t+1} + u_{t+2}$. The conditional moments are therefore *(a)* $\mathrm{E}_t\, r^e_{t+1} = a + z_t$ and $\mathrm{Var}_{t+1}(r^e_{t+1}) = \Sigma_{uu}$; *(b)* $\mathrm{E}_t\, r^e_{t+2} = a + \phi z_t, \mathrm{Var}_t(r^e_{t+2}) = \theta \Sigma_{uu} \theta' + \Sigma_{uu} + \Sigma_{\eta\eta}$; and *(c)* $\mathrm{Cov}_t(r^e_{t+1}, r^e_{t+2}) = \Sigma_{uu}\theta'$. Combining gives (13.11)–(13.12). □

13.3.4 Two-Period Investor (with Rebalancing)

We now consider the portfolio choice in t of an investor who will be able to rebalance in $t+1$. She solves a similar problem to that of an investor who cannot rebalance, where (13.9)–(13.10) define the relevant terms. However, in this case, $\mathrm{E}_t\, r_{p,t+2}$ and $\mathrm{Var}_t(r_{p,t+2})$ does not depend on the portfolio choice made in t (rather those in

$t+1$). In addition, $\mathrm{Cov}_t(r_{p,t+1}, r_{p,t+2})$ depends on both the portfolio choices in t (through $r_{p,t+1}$) and in $t+1$ (through $r_{p,t+2}$).

In principle, the covariance term is

$$\mathrm{Cov}_t(r_{p,t+1}, r_{p,t+2}) = v_t' \,\mathrm{Cov}_t(r_{t+1}^e, r_{t+2}^e v_{t+1}), \tag{13.13}$$

where v_t is the portfolio choice in t and v_{t+1} is the portfolio choice in $t+1$. The latter is a 1-period choice made in $t+1$, since that is the last time this 2-period investor makes an investment. Clearly, those weights are not known in t, so we apply an *approximation* by replacing v_{t+1} by its expected value obtained from the myopic case.

The optimal portfolio weights are as in (13.6), using

$$\mu^e = a + z_t + (1-\gamma)\Sigma_{uu}\theta' \,\mathrm{E}_t\, v_{t+1} \text{ and } \Sigma = \Sigma_{uu}. \tag{13.14}$$

(See below for a proof.) This version of "μ^e" captures both the expected return and the covariance term. The expression for $\mathrm{E}_t\, v_{t+1}$ is easily calculated as

$$\mathrm{E}_t\, v_{t+1} = \Sigma_{uu}^{-1}(a + \phi z_t + \mathrm{diag}(\Sigma_{uu})/2)/\gamma, \tag{13.15}$$

which follows directly from (13.6) and (13.8).

The solution (13.14)–(13.15) equals the myopic portfolio in two cases: *(1)* when $\gamma = 1$ (log utility); and/or when *(2)* $\theta = \mathbf{0}$ (no reversal or momentum).

See Figure 13.3 for an illustration of the case of a scalar risky return. The general pattern is similar to the case without rebalancing. This holds also for the multi-asset case in Figure 13.4.

Proof of (13.14). Since $r_{t+1}^e = a + z_t + u_{t+1}$ and $r_{t+2}^e = a + \phi z_t + \theta u_{t+1} + \eta_{t+1} + u_{t+2}$, the covariance is

$$\mathrm{Cov}_t(r_{p,t+1}, r_{p,t+2}) \approx v_{t+1}' \Sigma_{uu}\theta' \,\mathrm{E}_t\, v_{t+1}.$$

We also immediately get $\mathrm{E}_t\, r_{t+1}^e = a + z_t$ and $\mathrm{Var}_{t+1}(r_{t+1}^e) = \Sigma_{uu}$. Combining gives (13.14). □

13.3.5 Summary

The analysis in this section has shown that the optimal portfolio choice for a long-run (here, two-period) investor may differ substantially from that of a one-period investor *if* returns are non-iid. In particular, assets with long-run reversals are "safe" for a long run investor. This holds irrespective of whether the investor rebalances or not, although the mechanisms differ somewhat.

Bibliography

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

Back, K. E., 2010, *Asset Pricing and Portfolio Choice Theory*, Oxford University Press, Oxford.

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.

Blume, M. E., 1975, "Betas and Their Regression Tendencies," *Journal of Finance*, 30, 785–795.

Bondt, W. F. M. D., 1991, "What do economists know about the stock market?," *Journal of Portfolio Management*, 17, 84–91.

Bondt, W. F. M. D., and R. H. Thaler, 1990, "Do security analysts overreact?," *American Economic Review*, 80, 52–57.

Boni, L., and K. L. Womack, 2006, "Analysts, industries, and price momentum," *Journal of Financial and Quantitative Analysis*, 41, 85–109.

Brock, W., J. Lakonishok, and B. LeBaron, 1992, "Simple technical trading rules and the stochastic properties of stock returns," *Journal of Finance*, 47, 1731–1764.

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., 2005, *Asset pricing*, Princeton University Press, Princeton, New Jersey, revised edn.

Damodaran, A., 2012, *Investment Valuation: Tools and Techniques for Determining the Value of Any Asset*, John Wiley & Sons, Hoboken, NJ, 3rd edn.

Danthine, J.-P., and J. B. Donaldson, 2005, *Intermediate financial theory*, Elsevier Academic Press, 2nd edn.

Diebold, F. X., and R. S. Mariano, 1995, "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13, 253–265.

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.

Fabozzi, F. J., E. H. Neave, and G. Zhou, 2012, *Financial Economics*, John Wiley & Sons, Hoboken, NJ, 2nd 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.

Fama, E. F., and K. R. French, 2015, "A five-factor asset pricing model," *Journal of Financial Economics*, 116, 1–22.

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., 2018, *Econometric analysis*, Pearson Education Ltd, 8th edn.

Grinblatt, M., and S. Titman, 1993, "A Study of Monthly Mutual Fund Returns and Performance Evaluation Techniques," *Journal of Financial and Quantitative Analysis*, 28, 419–444.

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.

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.

JP Morgan, 1996, "RiskMetrics Technical Document," Discussion paper, JP Morgan, New York, NY.

Ledoit, O., and M. Wolf, 2003, "Improved estimation of the covariance matrix with an application to portfolio selection," *Journal of Empirical Finance*, 10, 603–621.

Ledoit, O., and M. Wolf, 2004, "Honey, I shrunk the sample covariance matrix," *Journal of Portfolio Management*, 30, 110–119.

Lintner, J., 1965, "The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets," *Review of Economics and Statistics*, 47, 13–37.

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.

Markowitz, H., 1952, "Portfolio Selection," *The Journal of Finance*, 7, 77–91.

Mayers, D., 1972, "Nonmarketable Assets, Market Segmentation, and the Level of Asset Prices," *Journal of Financial and Quantitative Analysis*, 7, 1–12.

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.

Mossin, J., 1966, "Equilibrium in a capital asset market," *Econometrica*, 34, 768–783.

MSCI Inc., 2024, *MSCI Global Equity Factor Model*.

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

Owen, J., and R. Rabinovitch, 1983, "On the Class of Elliptically Contoured Distributions with Mean–Variance Portfolio Evaluation," *Journal of Finance*, 38, 745–752.

Pennacchi, G., 2008, *Theory of Asset Pricing*, Pearson Education.

Rosenberg, B., and J. Guy, 1976, "Prediction of Beta from Accounting Data," *Financial Analysts Journal*, 32, 60–72.

Sharpe, W. F., 1964, "Capital asset prices: a theory of market equilibrium under conditions of risk," *Journal of Finance*, 19, 425–442.

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*.

Treynor, J. L., and K. Mazuy, 1966, "Can Mutual Funds Outguess the Market?," *Harvard Business Review*, 44, 131–136.