Portfolio optimization

In *portfolio optimization* (also known as *portfolio selection*), we invest in different assets, typically stocks, over some investment periods. The goal is to make investments so that the combined return on all our investments is consistently high. (We must accept the idea that for our average return to be high, we must tolerate some variation in the return, *i.e.*, some risk.) The idea of optimizing a portfolio of assets was proposed in 1953 by Harry Markowitz, who won the Nobel prize in economics for this work in 1990. In this section we will show that a version of this problem can be formulated and solved as a linearly constrained least squares problem.

Portfolio risk and return

Portfolio allocation weights. We allocate a total amount of money to be invested in n different assets. The allocation across the n assets is described by an allocation n-vector w, which satisfies $\mathbf{1}^T w = 1$, *i.e.*, its entries sum to one. If a total (dollar) amount V is to be invested in some period, then Vw_j is the amount invested in asset j. (This can be negative, meaning a short position of $|Vw_j|$ dollars on asset j.) The entries of w are called by various names including *fractional allocations*, *asset weights*, *asset allocations*, or just *weights*.

For example, the asset allocation $w = e_j$ means that we invest everything in asset j. (In this way, we can think of the individual assets as simple portfolios.) The asset allocation $w = (-0.2, 0.0, 1.2)$ means that we take a short position in asset 1 of one fifth of the total amount invested, and put the cash derived from the short position plus our initial amount to be invested into asset 3. We do not invest in asset 2 at all.

The *leverage* L of the portfolio is given by

$$L = |w_1| + \cdots + |w_n|,$$

the sum of the absolute values of the weights. If all entries of w are nonnegative (which is a called a *long-only portfolio*), we have $L = 1$; if some entries are negative, then $L > 1$. If a portfolio has a leverage of 5, it means that for every \$1 of portfolio value, we have \$3 of total long holdings, and \$2 of total short holdings. (Other definitions of leverage are used, for example, $(L-1)/2$.)

Multi-period investing with allocation weights. The investments are held for T periods of, say, one day each. (The periods could just as well be hours, weeks, or months). We describe the investment returns by the $T \times n$ matrix R, where R_{tj} is the fractional return of asset j in period t. Thus $R_{61} = 0.02$ means that asset 1 gained 2% in period 6, and $R_{82} = -0.03$ means that asset 2 lost 3%, over period 8. The jth column of R is the return time series for asset j; the tth row of R gives the returns of all assets in period t. It is often assumed that one of the assets is cash, which has a constant (positive) return μ^{rf}, where the superscript stands for *risk-free*. If the risk-free asset is asset n, then the last column of R is $\mu^{\mathrm{rf}}\mathbf{1}$.

Suppose we invest a total (positive) amount V_t at the beginning of period t, so we invest $V_t w_j$ in asset j. At the end of period t, the dollar value of asset j is $V_t w_j(1 + R_{tj})$, and the dollar value of the whole portfolio is

$$V_{t+1} = \sum_{j=1}^{n} V_t w_j (1 + R_{tj}) = V_t(1 + \tilde{r}_t^T w),$$

where $\tilde{r}_t^T$ is the tth row of R. We assume V_{t+1} is positive; if the total portfolio value becomes negative we say that the portfolio has *gone bust* and stop trading.

The total (fractional) return of the portfolio over period t, *i.e.*, its fractional increase in value, is

$$\frac{V_{t+1} - V_t}{V_t} = \frac{V_t(1 + \tilde{r}_t^T w) - V_t}{V_t} = \tilde{r}_t^T w.$$

Note that we invest the total portfolio value in each period according to the weights w. This entails buying and selling assets so that the dollar value fractions are once again given by w. This is called *re-balancing* the portfolio.

The portfolio return in each of the T periods can be expressed compactly using matrix-vector notation as

$$r = Rw,$$

where r is the T-vector of portfolio returns in the T periods, *i.e.*, the time series of portfolio returns. (Note that r is a T-vector, which represents the time series of total portfolio return, whereas $\tilde{r}_t$ is an n-vector, which gives the returns of the n assets in period t.) If asset n is risk-free, and we choose the allocation $w = e_n$, then $r = Re_n = \mu^{\mathrm{rf}}\mathbf{1}$, *i.e.*, we obtain a constant return in each period of μ^{rf}.

We can express the total portfolio value in period t as

$$V_t = V_1(1+r_1)(1+r_2)\cdots(1+r_{t-1}), \tag{17.1}$$

where V_1 is the total amount initially invested in period $t = 1$. This total value time series is often plotted using $V_1 = \$10000$ as the initial investment by convention. The product in (17.1) arises from re-investing our total portfolio value (including any past gains or losses) in each period. In the simple case when the last asset is risk-free and we choose $w = e_n$, the total value grows as $V_t = V_1(1+\mu^{\mathrm{rf}})^{t-1}$. This is called *compounded interest* at rate μ^{rf}.

When the returns r_t are small (say, a few percent), and T is not too big (say, a few hundred), we can approximate the product above using the sum or average of the returns. To do this we expand the product in (17.1) into a sum of terms, each of which involves a product of some of the returns. One term involves none of the returns, and is V_1. There are $t-1$ terms that involve just one return, which have the form $V_1 r_s$, for $s = 1, \ldots, t-1$. All other terms in the expanded product involve the product of at least two returns, and so can be neglected since we assume that the returns are small. This leads to the approximation

$$V_t \approx V_1 + V_1(r_1 + \cdots + r_{t-1}),$$

which for $t = T+1$ can be written as

$$V_{T+1} \approx V_1 + T\,\mathbf{avg}(r)V_1.$$

This approximation suggests that to maximize our total final portfolio value, we should seek high return, *i.e.*, a large value for $\mathbf{avg}(r)$.

Portfolio return and risk. The choice of weight vector w is judged by the resulting portfolio return time series $r = Rw$. The portfolio *mean return* (over the T periods), often shortened to just the *return*, is given by $\mathbf{avg}(r)$. The portfolio *risk* (over the T periods) is the standard deviation of portfolio return, $\mathbf{std}(r)$.

The quantities $\mathbf{avg}(r)$ and $\mathbf{std}(r)$ give the *per-period* return and risk. They are often converted to their equivalent values for one year, which are called the *annualized return and risk*, and reported as percentages. If there are P periods in one year, these are given by

$$P\,\mathbf{avg}(r), \qquad \sqrt{P}\,\mathbf{std}(r),$$

respectively. For example, suppose each period is one (trading) day. There are about 250 trading days in one year, so the annualized return and risk are given by $250\,\mathbf{avg}(r)$ and $15.81\,\mathbf{std}(r)$. Thus a daily return sequence r with per-period (daily) return 0.05% (0.0005) and risk 0.5% (0.005) has an annualized return and risk of 12.5% and 7.9%, respectively. (The squareroot of P in the risk annualization comes from the assumption that the fluctuations in the returns vary randomly and independently from period to period.)

Portfolio optimization

We want to choose w so that we achieve high return and low risk. This means that we seek portfolio returns r_t that are consistently high. This is an optimization problem with two objectives, return and risk. Since there are two objectives, there is a family of solutions, that trade off return and risk. For example, when the last asset is risk-free, the portfolio weight $w = e_n$ achieves zero risk (which is the smallest possible value), and return μ^{rf}. We will see that other choices of w can lead to higher return, but higher risk as well. Portfolio weights that minimize risk for a given level of return (or maximize return for a given level of risk) are called *Pareto optimal*. The risk and return of this family of weights are typically plotted on a risk-return plot, with risk on the horizontal axis and return on the vertical axis. Individual assets can be considered (very simple) portfolios, corresponding to $w = e_j$. In this case the corresponding portfolio return and risk are simply the return and risk of asset j (over the same T periods).

One approach is to fix the return of the portfolio to be some given value ρ, and minimize the risk over all portfolios that achieve the required return. Doing this for many values of ρ produces (different) portfolio allocation vectors that trade off risk and return. Requiring that the portfolio return be ρ can be expressed as

$$\mathbf{avg}(r) = (1/T)\mathbf{1}^T(Rw) = \mu^T w = \rho,$$

where $\mu = R^T\mathbf{1}/T$ is the n-vector of the average asset returns. This is a single linear equation in w. Assuming that it holds, we can express the square of the risk as

$$\mathbf{std}(r)^2 = (1/T)\|r - \mathbf{avg}(r)\mathbf{1}\|^2 = (1/T)\|r - \rho\mathbf{1}\|^2.$$

Thus to minimize risk (squared), with return value ρ, we must solve the linearly constrained least squares problem

$$\begin{array}{ll} \text{minimize} & \|Rw - \rho\mathbf{1}\|^2 \\ \text{subject to} & \begin{bmatrix} \mathbf{1}^T \\ \mu^T \end{bmatrix} w = \begin{bmatrix} 1 \\ \rho \end{bmatrix}. \end{array} \tag{17.2}$$

(We dropped the factor $1/T$ from the objective, which does not affect the solution.) This is a constrained least squares problem with two linear equality constraints. The first constraint sets the sum of the allocation weights to one, and the second requires that the mean portfolio return is ρ.

The portfolio optimization problem has the solution

$$\begin{bmatrix} w \\ z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 2R^TR & \mathbf{1} & \mu \\ \mathbf{1}^T & 0 & 0 \\ \mu^T & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 2\rho T\mu \\ 1 \\ \rho \end{bmatrix}, \tag{17.3}$$

where z_1 and z_2 are Lagrange multipliers for the equality constraints (which we don't care about).

As a historical note, the portfolio optimization problem (17.2) is not exactly the same as the one proposed by Markowitz. His formulation used a statistical model of returns, where instead we are using a set of actual (or *realized*) returns. (See exercise 17.2 for a formulation of the problem that is closer to the original formulation by Markowitz.)

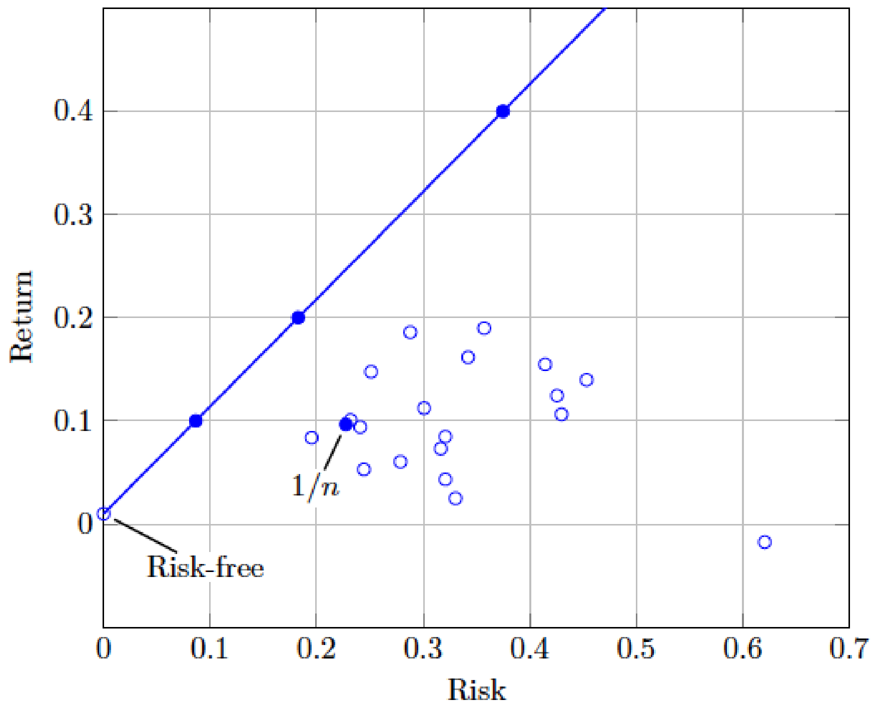

Figure 17.1 The open circles show annualized risk and return for 20 assets (19 stocks and one risk-free asset with a return of 1%). The solid line shows risk and return for the Pareto optimal portfolios. The dots show risk and return for three Pareto optimal portfolios with 10%, 20%, and 40% return, and the portfolio with weights $w_i = 1/n$.

Example

We use daily return data for 19 stocks over a period of 2000 days (8 years). After adding a risk-free asset with a 1% annual return, we obtain a 2000×20 return matrix R. The circles in figure 17.1 show the annualized risk and return for the 20 assets, *i.e.*, the points

$$\left[\begin{array}{c} \sqrt{250}\,\mathbf{std}(Re_i) \\ 250\,\mathbf{avg}(Re_i) \end{array}\right], \quad i = 1, \ldots, 20.$$

It also shows the Pareto-optimal risk-return curve, and the risk and return for the uniform portfolio with equal weights $w_i = 1/n$. The annualized risk, return, and leverage for five portfolios (the four Pareto-optimal portfolios indicated in the figure, and the $1/n$ portfolio) are given in table 17.1.

Figure 17.2 shows the total portfolio value (17.1) for the five portfolios. Figure 17.3 shows the portfolio values for a different test period of 500 days (two years).

	Return		Risk		
Portfolio	Train	Test	Train	Test	Leverage
Risk-free	0.01	0.01	0.00	0.00	1.00
10%	0.10	0.08	0.09	0.07	1.96
20%	0.20	0.15	0.18	0.15	3.03
40%	0.40	0.30	0.38	0.31	5.48
$1/n$	0.10	0.21	0.23	0.13	1.00

Table 17.1 Annualized risk, return, and leverage for five portfolios.

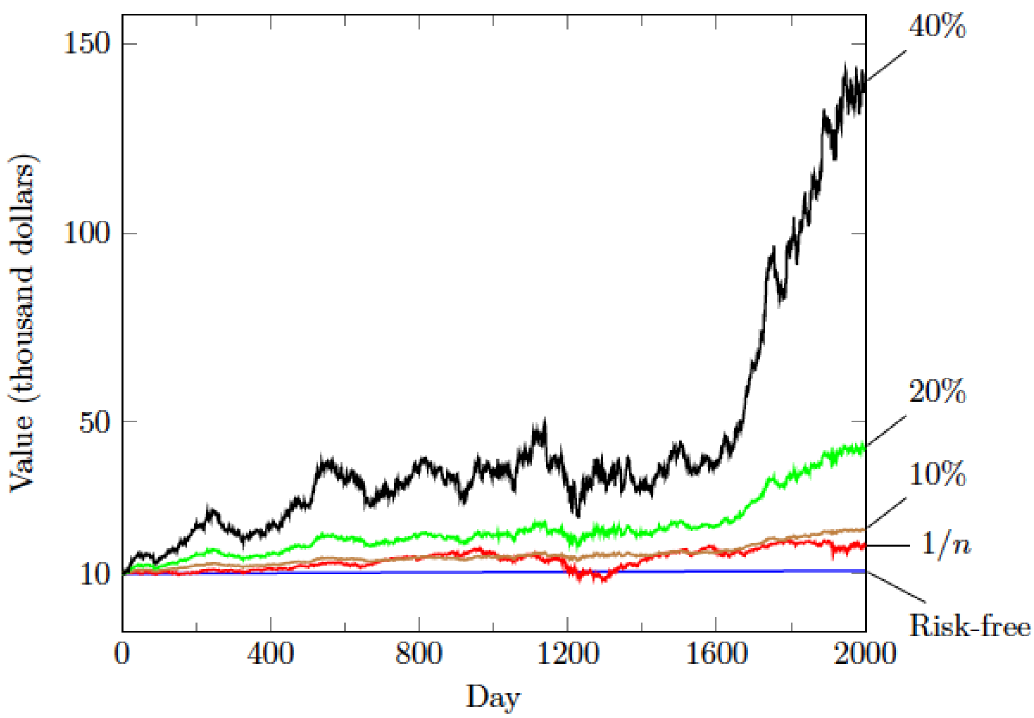

Figure 17.2 Total value over time for five portfolios: the risk-free portfolio with 1% annual return, the Pareto optimal portfolios with 10%, 20%, and 40% return, and the uniform portfolio. The total value is computed using the 2000×20 daily return matrix R.

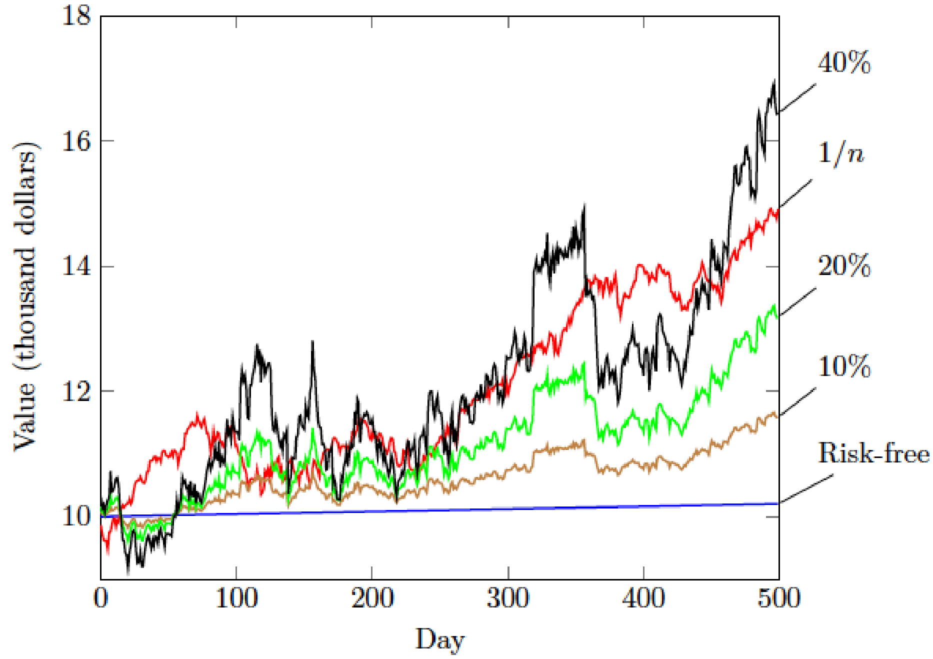

Figure 17.3 Value over time for the five portfolios in figure 17.2 over a test period of 500 days.

Variations

There are many variations on the basic portfolio optimization problem (17.2). We describe a few of them here; a few others are explored in the exercises.

Regularization. Just as in data fitting, our formulation of portfolio optimization can suffer from over-fit, which means that the chosen weights perform very well on past (realized) returns, but poorly on new (future) returns. Over-fit can be avoided or reduced by adding regularization, which here means to penalize investments in assets other than cash. (This is analogous to regularization in model fitting, where we penalize the size of the model coefficients, except for the coefficient associated with the constant feature.) One natural way to incorporate regularization in the portfolio optimization problem (17.2) is to add a positive multiple λ of the weighted sum of squares term

$$\sigma_1^2 w_1^2 + \cdots + \sigma_{n-1}^2 w_{n-1}^2$$

to the objective in (17.2). Note that we do not penalize w_n, which is the weight associated with the risk-free asset. The constants σ_i are the standard deviations of the (realized) returns, *i.e.*, $\sigma_i = \mathbf{std}(Re_i)$. This regularization penalizes weights associated with risky assets more than those associated with less risky assets. A good choice of λ can be found by back-testing.

Time-varying weights. Markets do shift, so it is not uncommon to periodically update or change the allocation weights that are used. In one extreme version of this, a new allocation vector is used in every period. The allocation weight for any period is obtained by solving the portfolio optimization problem over the preceding M periods. (This scheme can be modified to include testing periods as well.) The parameter M in this method would be chosen by validation on previous realized returns, *i.e.*, back-testing.

When the allocation weights are changed over time, we can add a (regularization) term of the form $\kappa\|w^{\mathrm{curr}} - w\|^2$ to the objective, where κ is a positive constant. Here w^{curr} is the currently used allocation, and w is the proposed new allocation vector. The additional regularization term encourages the new allocation vector to be near the current one. (When this is not the case, the portfolio will require excessive buying and selling of assets. This is called *turnover*, which leads to trading costs not included in our simple model.) The parameter κ would be chosen by back-testing, taking into account an approximation of trading cost.

Two-fund theorem

We can express the solution (17.3) of the portfolio optimization problem in the form

$$\begin{bmatrix} w \\ z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 2R^TR & \mathbf{1} & \mu \\ \mathbf{1}^T & 0 & 0 \\ \mu^T & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \rho \begin{bmatrix} 2R^TR & \mathbf{1} & \mu \\ \mathbf{1}^T & 0 & 0 \\ \mu^T & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 2T\mu \\ 0 \\ 1 \end{bmatrix}.$$

Taking the first n components of this, we obtain

$$w = w^0 + \rho v, \tag{17.5}$$

where w^0 and v are the first n components of the $(n+2)$-vectors

$$\begin{bmatrix} 2R^TR & \mathbf{1} & \mu \\ \mathbf{1}^T & 0 & 0 \\ \mu^T & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \begin{bmatrix} 2R^TR & \mathbf{1} & \mu \\ \mathbf{1}^T & 0 & 0 \\ \mu^T & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 2T\mu \\ 0 \\ 1 \end{bmatrix},$$

respectively. The equation (17.5) shows that the Pareto optimal portfolios form a *line* in weight space, parametrized by the required return ρ. The portfolio w^0 is a point on the line, and the vector v, which satisfies $\mathbf{1}^Tv = 0$, gives the direction of the line. This equation tells us that we do not need to solve the equation (17.3) for each value of ρ. We first compute w^0 and v (by factoring the matrix once and using two solve steps), and then form the optimal portfolio with return ρ as $w^0 + \rho v$.

Any point on a line can be expressed as an affine combination of two different points on the line. So if we find two different Pareto optimal portfolios, then we can express a general Pareto optimal portfolio as an affine combination of them. In other words, all Pareto optimal portfolios are affine combinations of just two portfolios (indeed, any two different Pareto optimal portfolios). This is the *two-fund theorem*. (*Fund* is another term for portfolio.)

Now suppose that the last asset is risk-free. The portfolio $w = e_n$ is Pareto optimal, since it achieves return μ^{rf} with zero risk. We then find one other Pareto optimal portfolio, for example, the one w^2 that achieves return $2\mu^{\mathrm{rf}}$, twice the risk-free return. (We could choose here any return other than μ^{rf}.) Then we can express the general Pareto optimal portfolio as

$$w = (1-\theta)e_n + \theta w^2,$$

where $\theta = \rho/\mu^{\mathrm{rf}} - 1$.