10.1. **Risk Budgeting.** We begin by recalling Euler's homogeneous function theorem.

*Theorem* 10.1 (Euler). Let $f(x_1, \ldots, x_k)$ be a smooth homogeneous function of degree $n$. That is,

$$f(tx_1, \ldots, tx_k) = t^n f(x_1, \ldots, x_k). \tag{10.1}$$

Then the following identity holds

$$x_1 \frac{\partial f}{\partial x_1} + \cdots + x_k \frac{\partial f}{\partial x_k} = nf.$$

*Proof.* By homogeneity, the relation (10.1) holds for all $t$. Taking the $t$-derivative of both sides, we establish that the following identity holds for all $t$:

$$x_1 \frac{\partial f}{\partial x_1}(tx_1, \ldots, tx_k) + \cdots + x_k \frac{\partial f}{\partial x_k}(tx_1, \ldots, tx_k) = nt^{n-1} f(x_1, \ldots, x_k).$$

To obtain the result of the theorem, it suffices to set $t = 1$ in the previous formula. □

Sometimes the differential operator

$$x_1 \frac{\partial}{\partial x_1} + \cdots + x_k \frac{\partial}{\partial x_k}$$

is called the *Euler operator*. An equivalent way to state the theorem is to say that homogeneous functions are eigenfunctions of the Euler operator, with the degree of homogeneity as the eigenvalue.

Let us consider a portfolio of $n$ assets. We define $x_i$ as the exposure of the $i$-th asset and $R(x)$ as a risk measure for the portfolio $x = (x_1, ..., x_n)$. Note that, at this level of generality, the $x_i$ could just as well be exposures to factors (such as style premia) instead of exposures to individual assets.

Assume $R(x)$ is homogeneous of degree 1, or in other words $R(tx) = tR(x)$. Then by Euler's theorem one has

return risk sources $r_i$

$$R(x) = \sum_{i=1}^{n} x_i \frac{\partial R}{\partial x_i}$$

$r = \sum_m X_m \; r_i$

exposures observable at time $t$

returns of risk sources over same period $[t, t+1]$

One says the risk measure is the sum of the product of exposure times marginal risk.

$x^T = h^T X$

Define the *risk contribution* of the $i$-th asset (or factor) as follows:

$$\mathrm{RC}_i(x) = x_i \frac{\partial R}{\partial x_i}$$

Again by Euler's theorem, this entails $R(x) = \sum_{i=1}^{n} \mathrm{RC}_i(x)$. Suppose we found some $x$ such that $\mathrm{RC}_i(x) = 1/n$ for all $i = 1, \ldots, n$. This would then mean that the total risk must be

$$R(x) = \sum_{i=1}^{n} \mathrm{RC}_i(x) = \sum_{i=1}^{n} \frac{1}{n} = 1.$$

These might actually be convenient units in which to work. Recall that $R(tx) = tR(x)$ for any $t > 0$, so if we've found $x$ such that $\mathrm{RC}_i(x) = 1/n$, it follows we can scale that $x$ to get a portfolio of any desired risk level, and the scaled portfolio will still have the "equal risk contribution" property that $\mathrm{RC}_i(x) = \mathrm{RC}_j(x)$ for all pairs $i, j$.

*Definition* 10.1. A *risk budgeting portfolio* is defined to be any portfolio whose exposures $x$ satisfy the following system of nonlinear equations:

$$\mathrm{RC}_i(x) = b_i \quad \forall\ i = 1, \ldots, n. \tag{10.2}$$

where $b_i > 0\ \forall\ i = 1, \ldots, n$ and $\sum_i b_i = 1$. A special case is the *equal risk contribution (ERC)* portfolio in which all $b_i = 1/n$.

The most commonly studied risk measure is volatility,

$$R(x) = \sigma(x) = (x'\Sigma x)^{1/2}$$

In this case, the marginal risk and the risk contribution of the $i$-th asset are respectively:

$$\frac{\partial R}{\partial x_i} = \frac{(\Sigma x)_i}{\sigma(x)}, \quad \text{and} \quad \mathrm{RC}_i = x_i \frac{(\Sigma x)_i}{\sigma(x)}.$$

Note that if the asset returns are Gaussian, the value-at-risk of the portfolio is:

$$\mathrm{VaR}(x; \alpha) = \Phi^{-1}(\alpha)\sigma(x)$$

where $\Phi$ is the CDF of the normal distribution. It is useful to memorize the most commonly-used values of $\Phi^{-1}$ for quick mental calculations:

$$\Phi^{-1}(0.95) \approx 1.64 \quad \text{and} \quad \Phi^{-1}(0.99) \approx 2.33 \tag{10.3}$$

So the 95% 1-day VaR for a normal distribution is approximately 1.6 times its daily volatility.

Similarly, the expected shortfall assuming a normal distribution is given by

哪来的？

$$\mathrm{ES}(x; \alpha) = \frac{1}{2\pi(1-\alpha)} \sigma(x) \exp\left[ -\frac{1}{2} \Phi^{-1}(\alpha)^2 \right] \tag{10.4}$$

Note that (10.3) and (10.4) are both proportional to $\sigma(x)$, so for normally-distributed returns, risk budgeting via $\sigma(x)$ could be trivially translated into budgeting VaR or ES by multiplying the risk budgets by suitable factors involving $\Phi^{-1}(\alpha)$. Hence in what follows, we mostly focus on $R(x) = \sigma(x)$.

$r_i = \beta \cdot r_M$



Let $\beta_i(x)$ denote the beta of asset $i$ to a portfolio $x$, ie

$$\beta_i(x) = \frac{\operatorname{cov}(r_i, x'r)}{\operatorname{var}(x'r)} = \frac{(\Sigma x)_i}{x'\Sigma x} = \frac{(\Sigma x)_i}{\sigma(x)^2}$$

Hence the risk contribution is

这个式子哪来的？

$$x_i\beta_i(x)\sigma(x) = x_i\frac{(\Sigma x)_i}{\sigma(x)} = \mathrm{RC}_i$$

If $x$ is a solution to the risk-budgeting problem (10.2), then $\mathrm{RC}_i(x) = b_i$ so we have for any pair $i, j$

$$\begin{aligned} x_i\beta_i(x)\sigma(x) &= b_i \\ x_j\beta_j(x)\sigma(x) &= b_j \end{aligned}$$

Divide the first equation by the second:

$$\frac{x_i\beta_i(x)}{x_j\beta_j(x)} = \frac{b_i}{b_j} \quad \Rightarrow \quad b_j x_i \beta_i(x) = b_i x_j \beta_j(x)$$

where again, the latter holds for all $i, j$ assuming $x$ is a solution to (10.2).

Moving things around,

$$b_j\beta_j(x)^{-1}x_i = b_i\beta_i(x)^{-1}\,x_j$$

Then sum both sides over $j$ to find

$$\frac{x_i}{\sum_i x_i} = \frac{b_i\beta_i(x)^{-1}}{\sum_j b_j\beta_j(x)^{-1}}$$

Circular 求β 需x 求x 需β.

$$x_i = \frac{b_i\beta_i(x)^{-1}}{\sum_j b_j\beta_j(x)^{-1}} \cdot \sum_i x_i$$

This does *not* constitute a solution to the problem, since $x$ appears on both sides, but it does help us understand the problem by telling us a property that the solutions must have at optimality: the weight allocated to the component $i$ is inversely proportional to its beta to the portfolio.

One can find risk-budgeting portfolios by the most classical method imaginable: least squares. Referring to (10.2), one can solve:

$$x^* = \underset{x}{\operatorname{argmin}} \sum_{i=1}^{n} (\mathrm{RC}_i(x) - b_i)^2 \,.$$

This is nonlinear least squares as $\mathrm{RC}_i(x)$ is typically a nonlinear function of $x$. The least-squares function is of course bounded below by zero, so the problem has at least one solution, but it is natural to wonder whether the solution is unique. This is answered by showing that the problem is equivalent to a convex optimization problem, as we now show.

Assume $\Sigma \succ 0$ is positive-definite, hence the following optimization problem is strictly convex:

把这个问题化为 standard form.

$f_1(y) = -\sum_i b_i \ln y_i + c \le 0.$

for typical convex optimization problem

$f_i(y) \le 0$ $\quad f_i$ is convex

$-f_i(y) \ge 0$ $\quad -f_i(y)$ is concave

$$\min \sigma(y) \quad \text{subject to} \quad \sum_i b_i \ln y_i \geq c \;\; \text{where:} \;\; \sigma(y) = (y'\Sigma y)^{1/2}$$

$\sum_i b_i \ln y_i$ is concave

so. 这是一个 valid 的 convex optimization 的条件

where $c$ is an arbitrary constant. Note that the domain of the problem $\mathcal{D}$

$$\mathcal{D} = \{y \in \mathbb{R}^n : y_i > 0 \;\forall\; i = 1, \ldots, n\}$$

because the constraint function isn't defined outside of this domain.

The Lagrangian is

$$L(y, \lambda) = \sigma(y) - \lambda\Big(\sum_i b_i \ln y_i - c\Big)$$

Note that the constraint will be active at optimality, because if we remove it, the solution would be $y^* = 0$ which isn't in the domain of the problem. Another, possibly more intuitive way to see this is as follows: start with any feasible point $y$. As we move $y$ increasingly close to the origin in the norm $\sigma(y)$, the components of $y$ become smaller, and $\ln y_i$ could be made arbitrarily negative, so by the intermediate value theorem eventually we hit $\sum_i b_i \ln y_i = c$. Hence at optimality $\lambda^* > 0$; we can then interpret the role of the arbitrary constant $c$: it serves to determine the Lagrange multiplier $\lambda^*$. We shall see that the value of $\lambda^*$ is arbitrary – it just rescales the portfolio.

y scale by t

这部分相当于分子分母同乘 t 所以不变.

$$\frac{\partial \sigma(y)}{\partial y_i} = \frac{(\Sigma y)_i}{\sigma(y)}$$

Consider the Lagrangian first-order condition:

$$(\nabla L)_i = \frac{\partial \sigma(y)}{\partial y_i} - \lambda \frac{b_i}{y_i} = 0$$

$$\Rightarrow y_i \frac{\partial \sigma(y)}{\partial y_i} = \lambda^* b_i$$

⇒ 找到 $y_i$, $\lambda^*$ 后, 可以 rescale $y_i$ 使得 $\lambda^* = 1$

This means we have shown that the risk contributions are proportional to the risk budgets. Hence (up to a scaling) $y$ solves the risk-budgeting problem.

10.2. **Relation to utility theory.** The von Neumann Morgenstern theorem implies that any *rational* investor must have a utility function, and such an investor decides between different lotteries by optimizing expected utility of wealth. Hence various investment schemes and theories and such must be regarded with suspicion unless they can be derived from utility theory somehow. Mean-variance optimization corresponds to expected utility maximization under the assumption that the multivariate distribution of asset returns is elliptical. It is natural to wonder whether the ERC portfolio is optimal under any reasonable assumptions.

Equal Risk Contribution

↳ 这个 portfolio 有什么好处？

*Theorem* 10.2. Let $R$ be the correlation matrix of asset returns. Suppose that

$$R\mathbf{1} = m \cdot \mathbf{1}$$

→ 1 是一个 eigen vector.

eg. $R = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$

where $\mathbf{1} = (1, 1, \ldots, 1)$ denotes a vector of ones. Suppose further that all assets have the same ex ante Sharpe ratio. Then the mean-variance optimal portfolio coincides with the risk-parity portfolio, and both have weights proportional to inverse volatility.

*Proof.* Let $S = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$, then the Sharpe ratio assumption is

$$S^{-1}\mu = \eta \cdot \mathbf{1}$$

for some constant $\eta$. By definition of "correlations" the asset return covariance matrix satisfies $\Sigma = SRS$ and

$$\Sigma^{-1} = S^{-1}R^{-1}S^{-1}.$$

The mean-variance optimal portfolio has weights proportional to

$$\Sigma^{-1}\mu = S^{-1}R^{-1}S^{-1}\mu = \eta S^{-1}R^{-1}\mathbf{1} = \eta m\, S^{-1}\mathbf{1}.$$

Hence the mean-variance portfolio is proportional to inverse-volatility weighting.

We now show that, under these hypotheses, the ERC portfolio is also inverse-vol weighted. The risk-contributions are $\mathrm{RC}_i(x) = x_i(\Sigma x)_i/\sigma(x)$. These contributions add to $\sigma(x)$, so the ERC condition is

$x = h^T X$

$$x_i \frac{(\Sigma x)_i}{\sigma^2(x)} = \frac{1}{n} \ \forall\, i = 1, \ldots, n \tag{10.5}$$

For the rest of the proof, we fix $x_i = \sigma_i^{-1}$ and check that, under the assumptions of the theorem, this $x$ satisfies (10.5).

Indeed, for this choice of $x$, we have $Sx = \mathbf{1}$ hence $RSx = m \cdot \mathbf{1}$, hence

$$SRSx = mS\mathbf{1} = m(\sigma_1, \ldots, \sigma_n)^T.$$

$(SRSx)_i = m \cdot \sigma_i$

$x_i (SRSx)_i = m\, \sigma_i \cdot \sigma_i^{-1} = m$

Finally, $x'SRSx = nm = \sigma^2(x)$. It follows that $x_i(\Sigma x)_i = m$ hence

$$x_i \frac{(\Sigma x)_i}{\sigma^2(x)} = \frac{m}{nm} = \frac{1}{n}$$

This completes the proof □.

For a (mean-variance) optimal portfolio, the ratio of the marginal excess return to the marginal risk is the same for all assets, and equals the ex ante Sharpe ratio of the portfolio:

$$\frac{\partial_i \mu(x)}{\partial_i \sigma(x)} = \mathrm{SR}(x) \quad \text{for all } i = 1 \ldots n.$$

为啥？

where $\partial_i \equiv \partial/\partial x_i$.

If your fund has alphas, use mean-variance optimization

if your fund does not have alphas, use assumption that $\mu$ is proportional to marginal risk (i.e. risk parity)



Letting $\mu = \mathbb{E}[r - r_{rf}] \in \mathbb{R}^n$ denote the expected excess returns, one has $\nabla\mu(x) =$ ? $\mu$ and $\nabla\sigma(x) = \Sigma x/\sigma(x)$. Therefore the above is equivalent to:

$$E(g_1) = SR(x) \cdot \frac{(\Sigma x)_1}{\sigma(x)}$$
$$E(g_2) = SR(x) \cdot \frac{(\Sigma x)_2}{\sigma(x)}$$
$$\vdots$$
$$E(g_n) = SR(x) \cdot \frac{(\Sigma x)_n}{\sigma(x)}$$

$$\mu = \mathrm{SR}(x)\frac{\Sigma x}{\sigma(x)} = \mathrm{SR}(x)\frac{\partial\sigma(x)}{\partial x_i}$$

In other words, given any risk-budgeting portfolio, we can derive the $\mu$ that makes it also mean-variance optimal by noting that the $\mu_i$ must be proportional to the marginal risks $\partial\sigma(x)/\partial x_i$.

Risk parity is often used to analyse a basic investment problem involving the optimal allocation to equities and bonds. Equity returns are far more volatile, so the ERC portfolio strongly down-weights equities.

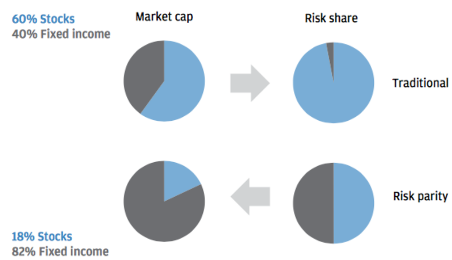


FIGURE 10.1. Traditional vs Risk Parity Allocation

The difference in weighting has a strong effect on the resulting performance.

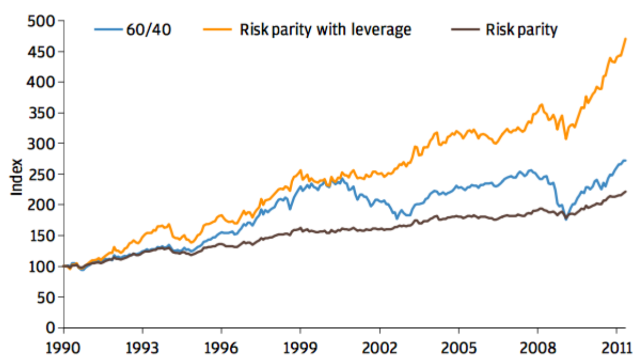


FIGURE 10.2. Performance of Traditional vs Risk Parity Allocation Since 1990

However, there are several important things going on here. If it happens that the Sharpe ratios are the same, then one point of view is that risk parity is performing well not because it's a good method, but because it's finding the mean-variance optimal portfolio by accident. It's also important to note that bonds have enjoyed a 30-year bull market which mostly encompasses the period over which we have reliable daily returns. Over such a period, any method which over-weights bonds relative to equities will realize a higher Sharpe ratio.

Lastly, let me say please be careful when reading the literature in this area. We are still mathematicians, even though we focus on an applied field of mathematics (mathematical finance). Mathematicians only make statements they can prove. Statements that are overly vague are the same as false statements. In doing a literature review for these notes, I encountered more than the usual amount of nonsense. For example, one paper stated:

> *The main difference between RB and MVO portfolios is that the last ones are based on optimization techniques. It implies that MVO portfolios are very sensitive to the inputs.*

— Bruder and Roncalli (2013)

From the same paper:

> *Mean-variance optimization, however, generally leads to portfolios concentrated in terms of weights.*

— Bruder and Roncalli (2013)

These statements are, of course nonsense or misleading. Mean-variance optimization generally leads to very well diversified portfolios if the return-generating process is of the APT form we have studied extensively in this class.

*Problem* 10.1. Complete the proof of Theorem 10.2, in other words show that, under the hypotheses of the theorem, the ERC portfolio is proportional to inverse-volatility weighting.