Smart Beta Strategies: Risk Budgeting and Parity

Risk Budgeting

- We begin by recalling Euler's homogeneous function theorem.

Theorem 7.1 (Euler)

Let $f(x_1, \dots, x_k)$ be a smooth homogeneous function of degree n on $\mathbb{R}^k$. That is, for all $t > 0$

$$f(tx_1, \dots, tx_k) = t^n f(x_1, \dots, x_k). \tag{7.1}$$

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 (7.1) holds for all $t > 0$. Taking the t-derivative of both sides, we establish that the following identity holds for all $t > 0$:

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

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

- 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.

- In fact, for the mathematically inclined, a much stronger result than theorem 7.1 is true: the theorem holds in the "if and only if" sense, and moreover, permits an extension to any open domain in $\mathbb{R}^k$.
- We leave the proof of these results as an exercise.

- 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

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

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

- 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 7.1

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

$$\mathrm{RC}_i(x) = b_i \;\; \forall \; i = 1, \ldots, n. \tag{7.2}$$

where $b_i > 0 \;\forall\; i = 1, \ldots, n$ and $\sum_i b_i = 1$. A common special case, which we will study in some detail, 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 Φ is the CDF of the normal distribution.
- It is useful to memorize the most commonly-used values of Φ^{-1} for quick mental calculations:

$$\Phi^{-1}(0.95) \approx 1.64 \quad \text{and} \quad \Phi^{-1}(0.99) \approx 2.33 \tag{7.3}$$

- 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

$$\mathrm{ES}(x;\alpha) = \frac{1}{2\pi(1-\alpha)}\sigma(x)\exp\Big[-\frac{1}{2}\Phi^{-1}(\alpha)^2\Big] \qquad (7.4)$$

- Note that (7.3) and (7.4) 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 $R(x) = \sigma(x)$.

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

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

- Hence the risk contribution is

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

- If x is a solution to the risk-budgeting problem (7.2), 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 (7.2).

- 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

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

- 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.

Solution by Convex Optimization

- One can find risk-budgeting portfolios by the most classical method imaginable: nonlinear least squares.
- Referring to (7.2), one can solve:

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

- 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 may be answered by showing that the problem is equivalent to a certain convex optimization problem, which under certain assumptions is strictly convex and has a unique minimum.

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

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

  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

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

- 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 λ^*.
- We shall see that the value of λ^* is arbitrary – it just rescales the portfolio.

- Consider the Lagrangian first-order condition:

$$
\begin{aligned}
(\nabla L)_i &= \frac{\partial \sigma(y)}{\partial y_i} - \lambda \frac{b_i}{y_i} = 0 \\
&\Rightarrow y_i \frac{\partial \sigma(y)}{\partial y_i} = \lambda^* b_i
\end{aligned}
$$

- Hence, at optimality, the risk contributions are proportional to the risk budgets.
- Hence (up to a scaling) y solves the risk-budgeting problem.

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 plans, schemes, and theories 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 there exist any reasonable assumptions consistent with utility theory, under which the ERC portfolio is the optimal one.

Theorem 7.2

Let R be the correlation matrix of asset returns. Suppose that

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

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 $S = \mathrm{diag}(\sigma_1, \dots, \sigma_n)$, then the Sharpe ratio assumption is
$$S^{-1}\mu = \eta \cdot 1$$
for some constant η.
- 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

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

- 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

$$x_i \frac{(\Sigma x)_i}{\sigma^2(x)} = \frac{1}{n} \ \forall \ i = 1, \ldots, n \tag{7.5}$$

- 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 (7.5).

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

$$SRSx = mS1 = m(\sigma_1, \ldots, \sigma_n)^T.$$

- 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 $\square$
- .

- 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:

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

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

- 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:

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

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

- Risk parity is often used to analyse a classic investment problem involving the optimal allocation to equities and bonds.

Equity returns are far more volatile, so the ERC portfolio strongly down-weights equities.

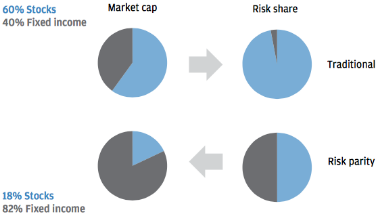

Figure: Traditional vs Risk Parity Allocation

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

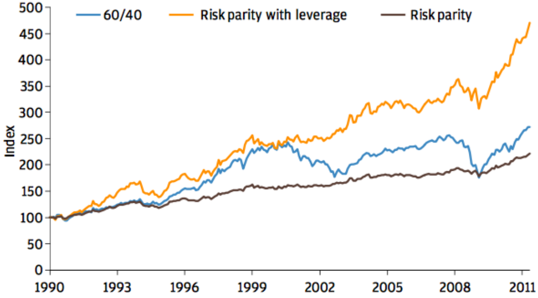

Figure: Performance of Traditional vs Risk Parity Allocation Since 1990

- However, there are several important things going on here.
- If it happens that the Sharpe ratios are the same, then perhaps risk parity is performing well not because it's a good method in general, 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 return data.
- Over such a period, any method which over-weights bonds relative to equities would have realized 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 as bad 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.