Lecture Notes - Performance-Seeking Portfolios

Lionel MARTELLINI

October 29, 2025

The risky portfolio should be efficient at harvesting risk premia across and within asset classes: It's all about diversification! If one could reliably predict future returns, there would be no need or desire diversification; one would instead invest in the highest-returning asset at each point in time. Crystal balls, however, hardly exist in the real world, and investors do not know in advance how prices will evolve even in the near future. In order to enjoy a higher expected return, some risk must be taken. The goal of diversification is precisely to find the most efficient way to harvest risk premia across and within risky assets – in other words, earn the highest expected return for a given risk budget, and this is achieved by diversifying away the largest possible amount of unrewarded risk.

1 The Holy Grail: The Maximum Sharpe Ratio (MSR) Portfolio

Mean-variance analysis provides an unambiguous definition for a well-diversified portfolio. The best diversified portfolio is the *maximum Sharpe ratio portfolio*, i.e., the portfolio that achieves the highest expected return in excess of the risk-free rate per unit of risk taken. This portfolio is determined by the risk and return characteristics of the assets at hand which need to be empirically estimated. From a practical standpoint, the investment process is often divided in two steps, where the first step is to decide how much to allocate *across* asset classes such as equities, bonds, commodities, etc., and the second step is to allocate *within* each class, by weighting appropriately the various individual securities. The concepts that we introduce here apply to both steps, but there are some practical differences between asset allocation and portfolio construction. The most visible is that the number of constituents is in general much larger when we allocate within a class, where the investor faces hundreds or thousands of individual securities. At the asset class level, the securities are typically aggregated into indices, and the number of constituents rarely exceeds a dozen or so. This will considerably facilitate the implementation task at both the parameter estimation stage and the portfolio optimization stage.

1.1 Optimal Portfolios in the Absence of a Risk-Free Asset: Derivation of the Efficient Frontier

We consider N risky assets (say N stocks, or N asset class indices) for which we use the following standard notation: μ_i is the expected return of asset i, σ_i is its volatility and ρ_{ij} is the correlation between assets i and j. The covariance σ_{ij} between assets i and j is the product of their correlation and their volatilities, $\rho_{ij}\sigma_i\sigma_j$. A portfolio is defined by a set of weights w_i, where w_i represents the percentage allocation to the risky asset i. The weights allocated to risky assets are chosen at the initial date and are collected in a vector $\mathbf{w}$. For a buy-and-hold portfolio (that is, a portfolio which is formed initially and not rebalanced until the terminal date), the portfolio expected return is the weighted sum of expected returns

$$\mu_p = \sum_{i=1}^{N} w_i\mu_i,$$

and its variance is the sum of variances weighted by the squared weights, plus a series of covariance terms:

$$\begin{aligned}
\sigma_p^2 &= \sum_{i,j=1}^{N} w_i w_j \rho_{ij}\sigma_i\sigma_j && (1)\\
&= \sum_{i=1}^{N} w_i^2\sigma_i^2 + \sum_{\substack{i,j=1\\ i\neq j}}^{N} w_i w_j \rho_{ij}\sigma_i\sigma_j && (2)\\
&= \sum_{i=1}^{N} w_i^2\sigma_i^2 + 2\sum_{\substack{i,j=1\\ i>j}}^{N} w_i w_j \rho_{ij}\sigma_i\sigma_j. && (3)
\end{aligned}$$

The presence of the covariance terms reflects an intuitive idea: individual risks add up if they are positively correlated, thereby increasing the risk of a basket of assets, or they partially cancel out if they are negatively correlated, thereby decreasing aggregate risk.

A much more compact expression for the expected return and the variance of a portfolio is achieved by introducing vector notation: $\boldsymbol{\mu}$ is the vector of expected returns, $\boldsymbol{\Sigma}$ is the *covariance matrix*, and $\mathbf{w}$ is the vector of weights in the risky assets. Let us first assume that the risk-free asset is not part of the investment universe, and consider the problem of optimal allocation to risky assets only. The portfolio expected return and volatility are given in matrix notation as:

$$\mu_p = \mathbf{w}'\boldsymbol{\mu} \tag{4}$$

$$\sigma_p = \sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}} \tag{5}$$

If we specialize for simplicity the analysis to a mean-variance setting, where investors express preferences over only the first two moments (that is mean and variance) of portfolio returns distributions, the only relevant characteristics of asset returns are expected returns, volatilities and pairwise correlations. Correlations matter because the risk of a portfolio not only depends on the standalone risk of every constituent, but also on the interactions between constituents. Mathematically, the mean-variance optimization problem can be stated as:

$$\underset{\mathbf{w}}{Max}\,\mathbf{w}'\boldsymbol{\mu} \text{ s.t. } \sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}} = \overline{\sigma} \text{ and } \mathbf{w}'\mathbf{e} = 1 \tag{6}$$

where $\mathbf{e}$ is a N-dimensional vector filled with ones, and $\overline{\sigma}$ is a target level of volatility. Letting $\overline{\sigma}$ vary allows one to obtain the set of all efficient portfolios, also known as the efficient frontier. One can also obtain the efficient frontier by solving the dual optimization program:

$$\underset{\mathbf{w}}{Min}\sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}} \text{ s.t. } \mathbf{w}'\boldsymbol{\mu} = \overline{\mu} \text{ and } \mathbf{w}'\mathbf{e} = 1 \tag{7}$$

where $\overline{\mu}$ is a target level of expected return.

Exercise 1 *Describe how you would set the minimum and maximum levels for the expected return target $\overline{\mu}$.*

1.2 Optimal Portfolios in the Presence of a Risk-Free Asset: Derivation of the MSR Portfolio

Let us assume now that the investment universe consists of N risky securities plus a risk-free asset. Here, the risk-free asset is an asset that pays \$1 with certainty at the terminal date: it can be thought of as a default-free pure discount bond with maturity date matching the investment horizon. Its return over the investment horizon, which is known in advance, is denoted with r. Returns on the risky securities over the investment period are unknown ex-ante, and they may have any probabilistic distributions.

We first show that the location of all portfolios $R_p = wR_A + (1-w)r$ defined as linear combinations of a portfolio A and the risk-free asset forms a straight line with an intercept equal to the risk-free rate and a slope equal to the Sharpe ratio of portfolio A defined as $\frac{\mu_A - r}{\sigma_A}$. To see this, we note that the expected return and volatility of these portfolios are given by:

$$\mu_p = w\mu_A + (1-w)r \tag{8}$$

$$\sigma_p = w\sigma_A \tag{9}$$

From these last two equations, we obtain

$$w = \frac{\sigma_p}{\sigma_A} \tag{10}$$

$$\mu_p = \frac{\sigma_p}{\sigma_A}\mu_A + (1 - \frac{\sigma_p}{\sigma_A})r = \frac{\mu_A - r}{\sigma_A}\sigma_p + r \tag{11}$$

It is then easy to see that in the presence of a risk-free asset, the efficient frontier is the straight line that with the highest slope. In other words, the only efficient portfolio composed with risky assets is the maximum Sharpe ratio portfolio, a.k.a. maximum reward-to-risk ratio portfolio, or also the tangency portfolio. The portfolio Sharpe ratio measures the ex-ante reward $\mu_p - r$ per unit of risk taken σ_p. Mathematically, it is given by

$$\lambda_p = \frac{\mu_p - r}{\sigma_p} = \frac{\mathbf{w}'\boldsymbol{\mu} - r}{\sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}} = \frac{\mathbf{w}'\left(\boldsymbol{\mu} - r\mathbf{e}\right)}{\sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}}, \tag{12}$$

where $\mathbf{e}$ is a N-dimensional vector filled with ones. We are interested in maximizing this quantity subject to the budget constraint, which states that weights should add up to one. This can be done via a numerical optimization procedure, but having an explicit expression is interesting from a theoretical standpoint because it shows how the composition of the maximum Sharpe ratio (MSR) portfolio depends on the characteristics of the underlying assets. It is also practically useful because numerical maximization in large universes can be time-consuming. It turns out that an exact mathematical formula exists when short sales are permitted. Then, it can be shown that the weights of the MSR portfolio in the risky assets are given by the following concise formula

$$\mathbf{w}_{MSR} = \frac{\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - r\mathbf{e}\right)}{\mathbf{e}'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - r\mathbf{e}\right)}, \tag{13}$$

where $\boldsymbol{\mu} - r\mathbf{e}$ is the vector of expected returns of assets in excess of the risk-free rate and the normalization factor $\mathbf{e}'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - r\mathbf{e}\right)$ in the denominator is the sum of the elements of the vector in the numerator, so that the weights add up to 100%.

Exercise 2 *Prove that the maximization of the portfolio Sharpe ratio λ_p gives the weights $\mathbf{w}_{MSR} = \frac{\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r\mathbf{e})}{\mathbf{e}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r\mathbf{e})}$.*

The inputs required to compute the MSR weights are the expected returns and the covariances. There are N expected returns, but covariances are in much greater quantity: there are $N[N+1]/2$ of them. Indeed, a covariance σ_{ij} is determined by a pair of indices i and j, and there are $N \times N$ of these pairs. Among these, there are N pairs with identical indices (e.g., $(1,1)$, $(2,2)$ and so on), which correspond to the variances. Hence, there $N^2 - N$ pairs with distinct indices. Since the covariance of i and j is the same as the covariance of j and i, there are $[N^2 - N]/2$ independent covariances between distinct assets. Adding the N variances, this amounts to a total of $[N^2 - N]/2 + N = N[N+1]/2$ independent covariances.

1.3 Estimation of the MSR Portfolio

Consider a set of N risky assets, for which we observe the returns over a sample period: let D be the sample size, $t = 1, ..., D$ be the sample dates, and $\mathbf{R}_t$ be the N observed returns at the end of period t. The sample estimators for the mean and covariance matrix are:

$$\widehat{\boldsymbol{\mu}} = \frac{1}{D}\sum_{t=1}^{D}\mathbf{R}_t, \tag{14}$$

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{D}\sum_{t=1}^{D}\left[\mathbf{R}_t - \widehat{\boldsymbol{\mu}}\right]\left[\mathbf{R}_t - \widehat{\boldsymbol{\mu}}\right]'. \tag{15}$$

Let us assume for simplicity that the vectors $\mathbf{R}_t$ are normally distributed with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and are independent in the probabilistic sense. Under these conditions, a standard statistical result says that the estimators $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ are independent from each other, and that they have the respective distributions

$$\widehat{\boldsymbol{\mu}} \sim N\left(\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{D}\right), \tag{16}$$

$$\widehat{\boldsymbol{\Sigma}} \sim \frac{1}{D}W\left(D-1, \boldsymbol{\Sigma}\right), \tag{17}$$

where N is the standard normal distribution and $W\left(D-1, \boldsymbol{\Sigma}\right)$ denotes the Wishart distribution with $D-1$ degrees of freedom and covariance matrix $\boldsymbol{\Sigma}$. The first of these results implies that $\widehat{\boldsymbol{\mu}}$ is an unbiased estimator of $\boldsymbol{\mu}$, and the properties of the Wishart distribution imply that the expectation of $\widehat{\boldsymbol{\Sigma}}$ equals $[D-1]\boldsymbol{\Sigma}/D$, so $\widehat{\boldsymbol{\Sigma}}$ is a biased estimator of $\boldsymbol{\Sigma}$, but the bias can easily be removed by normalizing by $D-1$ instead of D in the estimator: $\widehat{\boldsymbol{\Sigma}} = \frac{1}{D-1}\sum_{t=1}^{D}\left[\mathbf{R}_t - \widehat{\boldsymbol{\mu}}\right]\left[\mathbf{R}_t - \widehat{\boldsymbol{\mu}}\right]'$.

Note that the variance of $\widehat{\mu}_i$ equals σ_i^2/D, and thus shrinks to zero as the number of sample dates grows to infinity. More formally, the law of large numbers implies that $\widehat{\boldsymbol{\mu}}$ converges almost surely to the true mean as D becomes infinite. At finite distance/sample size ($D < \infty$), it is important to have a confidence interval for the true parameter values. The central limit theorem gives us such an interval: there is a 95% probability for the true value to fall in the range

$$\left[\widehat{\mu}_i - 1.96\frac{\sigma_i}{\sqrt{D}}\ ,\ \widehat{\mu}_i + 1.96\frac{\sigma_i}{\sqrt{D}}\right].$$

In practice, the true volatility σ_i is unknown, but it can be replaced by its sample counterpart to obtain an asymptotic confidence interval. When D grows, the radius of the interval decreases, so we get a more accurate estimate.

Remark 1 *Unreliability of expected return estimates. This gain in precision as a function of the increase in sample size is unfortunately illusory. Indeed, we are usually interested in the* annualized *expected return, given by $\widehat{\mu}_i/h$, where h is the annualization factor, equal to the number of years within one period, and the radius of the confidence interval for the annualized expected return is obtained by dividing the radius for μ_i by h. For instance, if the sample data is monthly, then $h = 1/12$. Noting that the number of years in the sample is $A = hD$, the confidence interval can be written as*

$$\frac{1.96\sigma_i}{h\sqrt{D}} = \frac{1.96\sigma_i}{h\sqrt{\frac{A}{h}}} = \frac{1.96}{\sqrt{A}} \times \frac{\sigma_i}{\sqrt{h}}. \tag{18}$$

The quantity $\sigma_i/\sqrt{h}$ is the annualized volatility. Hence, for a given annualized volatility, the precision of the sample estimator for expected returns depends only on the number of years in the sample. In particular, nothing is gained by increasing the frequency of the data (e.g., by moving from monthly to daily observations) as long as the number of years is constant. More importantly, the sample length required to arrive at a reasonable precision is excessive. Consider for instance a typical stock with a volatility of 20% per year. With 20 years of data, the radius of the confidence interval for the annualized expected return is $1.96/\sqrt{20} \times 0.2 = 8.8\%$. *If the sample expected return is 10% per year, this means that the true value can be located with a 95% confidence level somewhere between 1.2% and 18.8%: this range is so large that we hardly learn something about the true value. To have a radius of 2%, which is a reasonable but not extremely high precision if the true parameter is close to 10%, one needs at least* $[1.96 \times 0.2/0.02]^2 = 384$ *years of data! Such long samples are of course unavailable, and even if they were, it would be highly doubtful that the true expected return has remained constant for such a long time. As a conclusion, sample estimators are highly unreliable estimates of expected returns.*

Fortunately, the situation is somewhat better for risk parameter estimates. To see this, consider the expression for the variance of covariance estimates, which follows from the properties of the Wishart distribution:

$$\sigma\left(\widehat{\sigma}_{ij}\right)^2 = \frac{D-1}{D^2}\left[\sigma_{ij}^2 + \sigma_i^2\sigma_j^2\right]. \tag{19}$$

The same equation holds in terms of the annualized covariances $\sigma_{ij,ann} = h\sigma_{ij}$ and annualized volatilities $\sigma_{i,ann} = \sqrt{h}\sigma_i$:

$$\sigma\left(\widehat{\sigma}_{ij,ann}\right)^2 = \frac{D-1}{D^2}\left[\sigma_{ij,ann}^2 + \sigma_{i,ann}^2\sigma_{j,ann}^2\right] \tag{20}$$

$$= \left[\frac{h}{A} - \frac{h^2}{A^2}\right]\left[\sigma_{ij,ann}^2 + \sigma_{i,ann}^2\sigma_{j,ann}^2\right]. \tag{21}$$

Thus, the variance of the sample estimator depends not only on the number of years in the sample (A) but also on the frequency of the data ($1/h$). In particular, for a fixed number of years, one gets a less volatile estimator for each covariance by increasing the frequency of the data (i.e., decreasing h), which was not the case for expected returns. This is a fundamental difference between expected returns and variances: the sampling frequency has no effect on the precision of sample means, but covariances can be estimated with increasing precision by sampling data at a higher frequency. In this context, it makes sense to consider portfolio construction techniques that only rely upon the easier-to-estimate covariance matrix.

2 Implementable Proxies for the MSR Portfolio

Given that sample-based estimates being highly unreliable, one is left with a dependency on hopefully meaningful priors.

2.1 Scientific (Optimization-Based) Diversification: Global Minimum Variance Portfolios

Since sample-based expected return estimates are bound to be unreliable, and given that covariance parameters are somewhat easier to estimate, one could be tempted to use the only scientifically diversified portfolio that relies only on risk parameters, namely the GMV portfolio, with weights given by $\mathbf{w}_{GMV}^* = \frac{\mathbf{\Sigma}^{-1}\mathbf{e}}{\mathbf{e}'\mathbf{\Sigma}^{-1}\mathbf{e}}$, as an empirical proxy for the MSR.

Exercise 3 *Prove that the minimization of the portfolio variance σ_p^2 gives the weights $\mathbf{w}_{GMV} = \frac{\mathbf{\Sigma}^{-1}\mathbf{e}}{\mathbf{e}'\mathbf{\Sigma}^{-1}\mathbf{e}}$.*

2.1.1 Limits of GMV Portfolios

While asset managers are routinely packaging minimum variance portfolios as performance-seeking portfolios, this choice, however, is not as innocuous as it may seem: using a GMV portfolio as a proxy for the MSR portfolio actually implicitly boils down to assuming identical expected returns, which is not a highly reasonable prior! In practice, the GMV portfolio tends to underweight assets that have the largest contributions to volatility. This includes constituents with high volatilities, but also those that have large covariances with other assets, as can be seen by decomposing portfolio variance across variances and covariances. An asset that has large positive covariances with all other assets is likely to be assigned a negative weight so as to reduce the magnitude of the second term in the right-hand side of the above equation. The structure of the GMV portfolio simplifies in some special cases. For instance, if all correlations are zero, each asset weight is proportional to the inverse of its variance, which makes it clear that the portfolio overweights the least risky assets. Alternatively, if all correlations and volatilities are identical, then all assets are indistinguishable in terms of their contributions to portfolio risk. Thus, the optimization program has no reason to favor one asset over the others, and the GMV is an equally-weighted portfolio. In the general case of correlated assets, the matrix inversion must be performed numerically.

Ex-ante, the GMV portfolio is suboptimal with respect to the Sharpe ratio maximization criterion, except of course when all expected returns are equal, and the GMV portfolio carries in general some specific risk: the reason is that for the purpose of minimizing global risk, it does not matter how the risk of an asset is split across factors and residuals. Systematic risk and idiosyncratic risk are both sources of volatility which contribute exactly in the same way to aggregate risk. In other words, they are indistinguishable in the variance minimization program. This stands in contrast with the Sharpe ratio maximization exercise, where they do not play the same role, since only systematic risk is rewarded.

That the GMV portfolio contains unrewarded risk signals that it is not a well-diversified portfolio per se, but it still can be regarded as an implementable proxy for the unobservable MSR portfolio. Unfortunately, the influential paper of DeMiguel et al. (2009) (see **Reading: 1/N**) confirms that portfolios intended as proxies for the true MSR, such as the GMV portfolios, do not systematically display better out-of-sample Sharpe ratios than naive equally-weighted portfolios, even when improved estimates are used for the covariance matrix and expected return vectors. Overall, it appears that ignoring completely the differences across assets yields better results than trying to estimate their risk and return parameters. The apparent superiority of the equally-weighted portfolio in empirical horse races, in spite of its extreme simplicity, leads to question of whether naive diversification can prove helpful which is a subject we discuss next. Before we turn to the benefits of naive diversification, let us provide some information on how to improve scientifically diversified portfolios. Remember that the main problem with GMV portfolios is that the "magic" of diversification does not work properly for them: low volatility components are not penalized and therefore substantially over-weighted; minimum volatility does not just give up on estimating expected returns, it also gives up on diversification!

2.1.2 Improving GMV Portfolios

In this context, one possible improvement to the GMV portfolio consists in helping the optimizer focus on the covariance terms (where lies all the "magic" of diversification), as opposed to the variance terms. This can be done by forcing the assumption that all volatilities are equal $\sigma_i = \sigma$ for all i. As a result the variance minimization programs becomes:

$$\begin{aligned} Min\sigma_p^2 &= Min \sum_{i,j=1}^{N} w_i w_j \rho_{ij} \sigma_i \sigma_j && (22) \\ &= Min\sigma^2 \sum_{i,j=1}^{N} w_i w_j \rho_{ij} && (23) \\ &= Min \sum_{i,j=1}^{N} w_i w_j \rho_{ij} && (24) \end{aligned}$$

This portfolio is known as the *max decorrelation portfolio*, not a particularly pretty name for what is actually in general quite an efficient portfolio. Another related improvement consists in maximizing the so-called diversification index, defined as the weighted average volatility of the components of the portfolio to the portfolio volatility:

$$DI = \frac{\sum_{i=1}^{N} w_i \sigma_i}{\sqrt{\sum_{i,j=1}^{N} w_i w_j \sigma_{ij}}} \qquad (25)$$

The portfolio maximizing the diversification index is known as the *max diversification portfolio*, or the *most diversified portfolio* (Choueifaty and Coignard (2008)).[1] It is straightforward to check that the max diversification portfolio coincides with the MSR portfolio if and only if the excess return on each asset is proportional to its volatility: $\mu_i - r = \lambda\sigma_i$, which boils down to assuming that all assets have the same Sharpe ratio $\lambda = \frac{\mu_i - r}{\sigma_i}$. This assumption is at odds with the key insight from asset pricing models, such as Sharpe's (1964) Capital Asset Pricing Model (CAPM), which instead predicts that only systematic risk, as measured by a stock beta, and not total risk, as measured by a stock volatility, is rewarded. Assuming that the CAPM is the true asset pricing model, we then have: $\mu_i - r = (\mu_M - r)\beta_i$, implying that Treynor ratios, and not Sharpe ratios, are equal across stocks: $\frac{\mu_i - r}{\beta_i} = \mu_M - r$, a constant for all stocks. In these circumstances, the Sharpe ratio maximization program becomes

$$Max = \frac{\sum_{i=1}^{N} w_i (\mu_M - r)\beta_i}{\sqrt{\sum_{i,j=1}^{N} w_i w_j \sigma_{ij}}} = Max \frac{\sum_{i=1}^{N} w_i \beta_i}{\sqrt{\sum_{i,j=1}^{N} w_i w_j \sigma_{ij}}} = Max \frac{\beta_p}{\sigma_p}, \qquad (26)$$

which is yet another optimization-based proxy for the MSR portfolio that does not require sample-based estimates for expected returns.

2.2 Naive (Heuristic) Diversification: From EW Portfolios to Risk Parity Portfolios

As indicated above, the EW portfolio proves a surprisingly hard to beat benchmark. It turns out that one can regard the EW portfolio as the outcome of an optimization program, which is convenient whenever additional constraints have to be imposed (e.g., tracking error constraints).

Proposition 1 *The EW portfolio maximizes the effective number of constituents (ENC), which is defined as:*

$$ENC = \frac{1}{\sum_{i=1}^{N} w_i^2}. \qquad (27)$$

Proof. It is easy to see that the ENC achieves a minimum equal to 1 for a fully concentrated portfolio, and a maximum (for long-only portfolios) equal to N for an equally-weighted portfolio. ■

While it can be regarded as a natural starting point or benchmark, as a substitute to a CW portfolio, this

[1] Choueifaty, Yves, and Yves Coignard. 2008. Toward maximum diversification. The Journal of Portfolio Management 35: 40-51.

maxENC/EW/naively diversified portfolio, however, makes no use of the information contained in empirical estimates for risk parameters, which are quite reliable. As a result, naive diversification can lead to high risk concentration when applied to assets with non homogenous risks. To see this, let us consider a portfolio invested 50% in a 1% vol bond and 50% in (uncorrelated) 30% vol stock. For this portfolio, weights are highly diversified, but risk is highly concentrated. This is due to differences in volatility levels:

$$(50\%)^2(30\%)^2 >> (50\%)^2(1\%)^2 \tag{28}$$

A risk budgeting approach would suggest to measure the dispersion of risk as opposed to $ contributions:

$$\begin{aligned} p_1 &= (50\%)^2(30\%)^2/\left[(50\%)^2(30\%)^2 + (50\%)^2(1\%)^2\right] = 0.11\% && (29) \\ p_2 &= (50\%)^2(1\%)^2/\left[(50\%)^2(30\%)^2 + (50\%)^2(1\%)^2\right] = 99.89\% && (30) \end{aligned}$$

In other words, the prescription is that naive diversification should apply to risk contributions as opposed to dollar contributions, that is choose the dollar contributions w_i so that the risk contribution w_i are equal, and equal to $1/N$. This improved implementation of naive diversification takes the form of another portfolio that can be estimated solely on the basis of the covariance matrix. This portfolio is the *risk parity* portfolio, defined to be such that all assets have the same contribution to the portfolio risk. To motivate an interest in this portfolio, let us first examine the quantitative measure of contribution of an asset to portfolio risk. In general, that is when correlations are not zero, the inverse covariance matrix $\mathbf{\Sigma}^{-1}$ has a complex structure, and an asset weight does not only depend on its own risk and performance characteristics but it is also a function of its correlations with the other constituents. Indeed, an asset contributes to portfolio risk through its variance, but also through its covariances with the other constituents. Intuitively, including an asset in the portfolio increases aggregate risk if this asset is highly volatile, but also if it covaries positively with other constituents, because fluctuations in other asset prices tend to be associated with fluctuations in the same direction (gain or loss) in this asset. Mathematically, the contribution of an asset to portfolio risk can be quantitatively defined by writing the portfolio variance as

$$\sigma_p^2 = \sum_{i,j=1}^{N} w_i w_j \sigma_{ij} = \sum_{i=1}^{N} w_i \left[\sum_{j=1}^{N} w_j \sigma_{ij}\right], \tag{31}$$

where σ_{ij} is the covariance of assets i and j. (Recall that the covariance σ_{ii} equals σ_i^2, the variance of asset i.) The term $w_i \sum_{j=1}^{N} w_j \sigma_{ij}$ is called the contribution of asset i to portfolio variance, and it clearly depends on σ_{ij} for all j. For example, for $N = 3$, we have:

$$\begin{aligned} \sigma_p^2 &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + w_3^2\sigma_3^2 \\ &\quad +2w_1w_2\sigma_{12} + 2w_1w_3\sigma_{13} + 2w_2w_3\sigma_{23} && (32) \\ &= w_1^2\sigma_1^2 + w_1w_2\sigma_{12} + w_1w_3\sigma_{13} \\ &\quad +w_2^2\sigma_2^2 + w_2w_1\sigma_{12} + w_2w_3\sigma_{23} \\ &\quad +w_3^2\sigma_3^2 + w_3w_1\sigma_{13} + w_3w_2\sigma_{23} && (33) \end{aligned}$$

Next, divide both sides by σ_p to arrive at portfolio volatility:

$$\sigma_p = \sum_{i=1}^{N} \frac{w_i}{\sigma_p}\left[\sum_{j=1}^{N} w_j \sigma_{ij}\right]. \tag{34}$$

This equation expresses portfolio volatility as the sum of contributions from the various assets. The contribution of asset i to volatility is formally defined as

$$c_i = \frac{w_i}{\sigma_p}\left[\sum_{j=1}^{N} w_j \sigma_{ij}\right]. \tag{35}$$

The biggest contributors are assets with large volatilities and/or large covariances with the others. Indeed, a highly volatile asset is obviously a big contributor to risk, and an asset that covaries positively with the other constituents also contributes to risk because shocks to the various assets add up. A second expression for the contribution of an asset is in terms of the covariance between the asset return (R_i) and the portfolio return (R_p):

$$c_i = \frac{w_i}{\sigma_p} cov\left[R_i, R_p\right]. \tag{36}$$

This formula makes it clear that the asset contribution to portfolio risk has the same sign as the covariance between the asset return and the portfolio return. A third expression for the contribution involves the sensitivity of portfolio volatility with respect to the allocation to asset i. Mathematically, $\sum_{j=1}^{N} w_j \sigma_{ij}$ is related to the partial derivative of portfolio volatility with respect to w_i, that is the change in portfolio variance for an infinitesimal change of w_i, other weights remaining fixed:

$$\begin{aligned} \frac{\partial \sigma_p}{\partial w_i} &= \frac{\partial\left(\sum_{i=1}^{N} w_i^2\sigma_i^2 + 2\sum_{i>j} w_i w_j \sigma_{ij}\right)^{1/2}}{\partial w_i} && (37) \\ &= \frac{1}{2}\left(2\sum_{j=1}^{N} w_j\sigma_{ij}\right)\left(\sum_{i=1}^{N} w_i^2\sigma_i^2 + 2\sum_{i>j} w_i w_j \sigma_{ij}\right)^{-1/2} && (38) \\ &= \frac{\sum_{j=1}^{N} w_j \sigma_{ij}}{\sigma_p} && (39) \end{aligned}$$

The contribution of asset i can thus be rewritten as

$$c_i = \frac{w_i}{\sigma_p}\left[\sum_{j=1}^{N} w_j \sigma_{ij}\right] = w_i \frac{\partial \sigma_p}{\partial w_i}. \tag{40}$$

The biggest contributors are the assets for which the allocation has a large impact on portfolio volatility. Asset contributions add up to the portfolio volatility:

$$\sigma_p = \sum_{i=1}^{N} w_i \frac{\partial \sigma_p}{\partial w_i}. \tag{41}$$

This equation is known as *Euler identity*. In fact, it holds not only for volatility, but also for any risk measure that is "homogenous of degree 1 in the weights".[2] This allows us to define contributions for other risk measures than volatility, as long as they satisfy the homogeneity condition (which is the case for instance for semi-volatility or Value-at-Risk).

A number of numerical routines can be used to find the risk parity portfolio, also known as or *equal risk contribution* (ERC) portfolio, defined as the portfolio for which the contribution to risk is identical for all assets; in other words it is a maxENC portfolio, except that here the ENC is applied to risk contributions (c_i), as opposed to dollar contributions (w_i), as in the EW portfolio. This ERC weighting scheme can be viewed as a response to the problem of excessive risk concentration in equally-weighted portfolios. It is the exact counterpart of the equally-weighted allocation with a different definition for eggs and baskets: baskets are still assets, but eggs are risk contributions as opposed to dollar contributions, so that the highest extent of diversification is achieved by spreading risk contributions evenly across assets.

Exercise 4 *In some specific cases, the risk parity portfolios can be directly computed. In particular, if all assets are uncorrelated from each other, one can show that the weights of the risk parity portfolio are inversely proportional to volatilities, leading to a portfolio known as the* inverse volatility portfolio. *To see this, first compute the portfolio variance in the uncorrelated case:*

$$\sigma_p^2 = \sum_{i=1}^{N} w_i^2 \sigma_i^2, \tag{42}$$

so the risk contribution of asset i is

$$c_{rel,i} = \frac{[w_i \sigma_i]^2}{\sigma_p^2}, \tag{43}$$

[2]Homogeneity of degree 1 means that if all weights are multiplied by a positive constant (which amounts to diluting the portfolio with cash or to adding leverage), the function is multiplied by the same constant.

Risk contributions are identical across constituents if

$$\frac{[w_i \sigma_i]^2}{\sigma_p^2} = \frac{1}{N} \Longrightarrow w_i^2 \sigma_i^2 = \frac{\sigma_p^2}{N} \tag{44}$$

$$\Longrightarrow w_i = \frac{1}{\sigma_i}\frac{\sigma_p}{\sqrt{N}} = \frac{\frac{1}{\sigma_i}}{\sum_{j=1}^{N} \frac{1}{\sigma_j}} \tag{45}$$

Remarkably, this property still holds in the presence of non-zero correlations across assets provided all correlations are equal.

With a more general correlation structure across assets, and possibly different correlations, the inverse volatility portfolio does not achieve risk parity, and no analytical expression is available for the weights. In this situation, numerical routines have to be employed. The RP portfolio is not an optimal portfolio according to mean-variance theory (in general it does not lie on the ex-ante efficient frontier), but it may prove ex-post better in terms of performance or risk-adjusted performance (Sharpe ratio) than proxies for efficient portfolios because it is not plagued by the errors in expected return estimates that dramatically affect these proxies. However, the RP portfolio can be related to mean-variance theory under some restrictive conditions on parameter values: it can be shown that risk parity weights maximizes the Sharpe ratio when all assets have the same Sharpe ratio and the same pairwise correlations.

2.3 Mixing Scientific and Heuristic Diversification

Scientific and heuristic approaches might be regarded as competing methods for constructing a well-diversified portfolio, but they can in fact be mixed for even better results. In a nutshell, the idea is to shrink the scientifically diversified weights towards the equal weighted or equal risk weighted portfolio.

2.3.1 Shrinkage Towards the Naive Portfolio with Hard Constraints

Consider first a very simple and common constraint, which states that all weights should be nonnegative (and must consequently be less than 100%). The primary reason for imposing this condition is that short sales are hardly feasible for most investors, either because buying on margin is costly or simply because short sales are prohibited. Institutional or regulatory reasons may also impose tighter bounds on weights, e.g. by restricting the percentage allocation to a given security or asset class to be less than a cap. Capping weights is also a means to ensure that at least a minimum number of constituents will have *positive* weights: for instance, if all weights are required to be less than 5%, then at least $1/0.05 = 20$ constituents will be assigned a positive weight. More generally, by taking the upper bound to be of the form δ/N, where N is the nominal number of constituents and δ is some number between 1 and N, then a portfolio with weights comprised between 0 and δ/N has at least N/δ non-zero weights: in other words, a percentage $1/\delta$ of the securities are effectively included in the portfolio.

It turns out that imposing bounds on weights in a variance minimization program allows to reduce sampling errors in the covariance matrix. This result is formally established by Jagannathan and Ma (2003).[3] The main argument goes as follows. Minimizing variance subject to a no short-sales constraint is equivalent to minimizing variance without any weight constraint, but with a modified covariance matrix, in which some of the covariances have been shrunk downwards. The modified covariances are those that involve at least one asset for which the constraint is "binding", that is one asset for which the constraint would be violated if it was not imposed. They tend to be large covariances in the original matrix, because assets that have large covariances with the others are the biggest contributors to portfolio risk, i.e. the least desirable ones in a minimum variance allocation. When there is a strongly dominant factor, like the market factor for stocks or the interest rate level for bonds, the constraint is more likely to be binding for high beta securities (long duration bonds in a fixed-income universe). In practice, the covariance matrix has to be estimated, but large covariances are more likely to be overestimated, so the process of shrinking covariances downwards can be regarded as an attempt to correct for the effects of overestimation. In a nutshell, it might be wrong to impose no short-sales constraints if parameters were perfectly known because the true global minimum variance (GMV) portfolio may contain short positions, but this may be justified in the presence of parameter uncertainty as a means to mitigate the effects of estimation errors. Empirically, Jagannathan and Ma (2003) show that when nonnegativity constraints are imposed, it makes little difference whether the sample covariance matrix or a more sophisticated estimator is used, and whether the sample data is monthly or daily: once the constraints are imposed, all estimators and both sampling frequencies produce roughly the same out-of-sample volatility. However, one problem with nonnegativity constraints in the presence of a dominant factor is that they are binding for many securities, resulting in portfolios concentrated in a few stocks. As explained above, imposing a cap mitigates this problem, but a large cap is needed to spread the portfolio across many securities: to ensure that at least one third of securities have positive weights in a 100-stock universe, all weights should be between 0 and 3%, which leaves little room for optimization.

[3] Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. The Journal of Finance, 58(4), 1651-1683.

2.3.2 Shrinkage Towards the Naive Portfolio with Soft (a.k.a. Norm) Constraints

DeMiguel et al. (2009)[4] pursue the idea of imposing weight constraints, by introducing a more global form of constraints, which they call "norm constraints". A general presentation of norm constraints is beyond the scope of these lecture notes, but two special cases stand out. The first consists in setting a cap on the amount of leverage, i.e. in setting an upper bound on the sum of negative weights: this nests the no-short sales constraint as a limit case, but allows for a controlled amount of leverage to be taken in case the constraint is relaxed. The second special case is when the sum of squared weights is capped to a maximum. By definition of the effective number of constituents, imposing that the sum of squared weights is capped to a maximum is equivalent to requiring a minimum ENC for the portfolio. In other words, the constrained MSR or GMV portfolio is a scientifically diversified portfolio subject to a heuristic diversification constraint. DeMiguel et al. (2009) show that this approach is equivalent to a specific form of "shrinkage" of the sample covariance matrix: to shrink the sample covariance matrix means to replace it with a weighted sum of itself and another matrix, known as the shrinkage target. This is done to achieve a balance between specification risk – which is null for the sample covariance matrix and potentially large for the target – and sample risk – which is large for the sample matrix but much lower for the target. In details, the minimum ENC constraint is equivalent to a shrinkage of the sample covariance matrix towards a target matrix with zero covariances and the same variance for each asset. This result makes intuitive sense: the minimum ENC constraint is meant to reduce the distance between the estimated GMV portfolio and the equally-weighted one, which would precisely be the true GMV portfolio if assets were uncorrelated with equal risk. Empirically, GMV portfolios with ENC constraints have often lower variance and higher Sharpe ratios than GMV portfolios with no short-sales constraints.

[4] DeMiguel, V., Garlappi, L., Nogales, F. J., & Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. Management science, 55(5), 798-812.