Sensitivity of Mean–Variance Models to Input Estimation

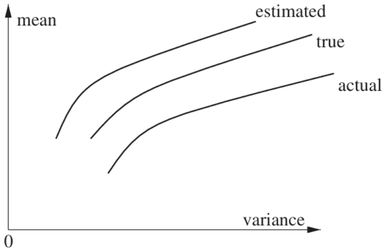
Sirong Luo

Faculty of Statistics and Data Sciences
Shanghai University of Finance and Economics

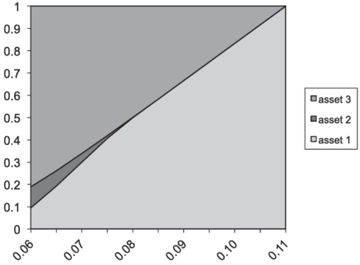
Introduction Example

Consider a simple portfolio optimization problem with three assets whose expected returns and covariance matrix are

$$\boldsymbol{\mu} = \begin{bmatrix} 0.11 \\ 0.10 \\ 0.05 \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} 0.250 & 0.225 & 0.045 \\ 0.225 & 0.250 & 0.045 \\ 0.045 & 0.045 & 0.090 \end{bmatrix} \tag{7.1}$$



(a) Efficient frontiers



(b) Area chart of long-only efficient portfolios for $\boldsymbol{\mu}$ and $\mathbf{V}$ as in (7.1)

Figure 1: the sensitivity of mean–variance models to the quality of the inputs

Figure 1(b) displays the composition of long-only efficient portfolios for this problem.

The picture makes sense from the pure optimization standpoint: assets 1 and 2 are similar but the expected return of asset 1 is slightly larger. Hence for higher target expected returns, the efficient portfolios have a much larger holding in asset 1 than in asset 2. However, from the portfolio construction standpoint this is unintuitive: assets 1 and 2 are very similar and, for all practical purposes, exchangeable because the slight difference could easily be due to estimation error.

Therefore, it would be more intuitive for the positions of these two assets to be roughly the same. We can also look at the problem in a different way: Suppose the expected returns of assets 1 and 2 were slightly perturbed so that they are swapped. Then the composition of the efficient portfolios would change drastically. This again is a fairly counterintuitive and unnatural behavior.

The input sensitivity of mean-variance models is a central issue in portfolio management and has been a subject of intense study. There is a tremendous upside potential in finding appropriate ways of harnessing the power of portfolio optimization without getting caught on this major shortcoming. We will next describe some of the most popular techniques that aim at mitigating this problem. The techniques can be classified in two main categories.

The first category of techniques tries to improve the quality of the inputs to the portfolio optimization problem. The second category of techniques aims to tweak the optimization procedure. One of the most widely used techniques in the first category is the BlackLitterman model introduced by Fisher Black and Bob Litterman at Goldman Sachs Asset Management. We will discuss this technique in some detail. We will also briefly describe another related technique based on Bayesian adjustments. We will subsequently discuss some techniques in the second category, namely resampled efficiency and robust optimization.

Black–Litterman Model

The basic idea of the Black-Litterman model is to tilt the market equilibrium returns to incorporate an investor's views. In principle a classical mean-variance model requires estimates of expected returns for all assets in the investment universe considered. This is typically an enormous task. Investment managers are unlikely to have detailed knowledge of all securities at their disposal. Typically, they have a specific area of expertise. Furthermore, some modern trading strategies are associated not with absolute but with relative rankings of securities. For instance, a pairs trading strategy corresponds to a forecast that one stock will outperform another one. The key insight of Black and Litterman was that there is a suitable way of combining the investor's views with the market equilibrium. The exposition of the Black-Litterman model below is based on Black and Litterman (1992), Fabozzi et al. (2007), and Litterman (2003).

Basic Assumption and Starting Point

The Black-Litterman model is an equilibrium-based model, meaning that the expected returns of the assets should be consistent with the market equilibrium unless the investor has some specific views. In other words, an investor without any views on the market should hold the market. We shall let $\boldsymbol{\pi}$ denote the equilibrium return vector, and $\mathbf{V}$ the covariance matrix of the asset returns.

The true expected return vector $\boldsymbol{\mu}$ is unknown. As a starting point, we assume that the equilibrium return vector serves as a reasonable prior estimate of the true return vector in the sense that

$$\boldsymbol{\mu} \sim N(\boldsymbol{\pi}, \mathbf{Q}).$$

That is, $\boldsymbol{\mu}$ is a multi-normal random vector with expected value $\boldsymbol{\pi}$ and covariance matrix $\mathbf{Q}$. The matrix $\mathbf{Q}$ represents the confidence on the equilibrium returns as an estimate of expected returns.

Expressing Investors' Views

A key ingredient of the Black-Litterman model is to incorporate investors' views on the expected returns. The framework is fairly flexible. An investor may have a few different views, each of them involving either a single asset (an absolute view) or several assets (a relative view). Formally, a collection of views is expressed as

$$\mathbf{P}\boldsymbol{\mu} = \mathbf{q} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Omega}).$$

Each row in the equation $\mathbf{P}\boldsymbol{\mu} = \mathbf{q} + \boldsymbol{\epsilon}$ is a view that represents a forecast. The term $\boldsymbol{\epsilon}$ represents the degree of confidence in the views. The covariance matrix $\boldsymbol{\Omega}$ is typically a diagonal matrix. A weak view is a view with large variance; a strong view is a view with small variance. In the extreme, a certain view is a view with zero variance. Each of the views can be absolute or relative as described above. For a concrete example, consider an asset allocation problem with seven asset classes: Australia, Canada, France, Germany, Japan, United Kingdom, United States. Suppose we have two views: - Return on Germany will be 12%. - UK will outperform US by 2%.

These views can be expressed as

$$\mu_4 = 12\% + \epsilon_1$$
$$\mu_6 - \mu_7 = 2\% + \epsilon_2.$$

In matrix notation this corresponds to $\mathbf{P}\boldsymbol{\mu} = \mathbf{q} + \boldsymbol{\epsilon}$ for

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \quad \mathbf{q} = \begin{bmatrix} 12\% \\ 2\% \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}.$$

The key insight of the Black-Litterman model is a proper way to combine the investor's views with the prior market equilibrium. First, consider the simpler case when the views are assumed to be certain; that is, $\boldsymbol{\Omega} = \mathbf{0}$ or equivalently the vector of expected returns must be tilted to satisfy the views $\mathbf{P}\boldsymbol{\mu} = \mathbf{q}$. In this case the posterior estimate of $\boldsymbol{\mu}$ given the prior $\boldsymbol{\mu} \sim N(\boldsymbol{\pi}, \mathbf{Q})$ and the views $\mathbf{P}\boldsymbol{\mu} = \mathbf{q}$ is

$$\hat{\boldsymbol{\mu}} = \boldsymbol{\pi} + \mathbf{Q}\mathbf{P}^\top \left(\mathbf{P}\mathbf{Q}\mathbf{P}^\top\right)^{-1} (\mathbf{q} - \mathbf{P}\boldsymbol{\pi}). \tag{7.2}$$

Some matrix algebra shows that indeed $\mathbf{P}\hat{\boldsymbol{\mu}} = \mathbf{q}$. That is, the posterior estimate satisfies the views $\mathbf{P}\boldsymbol{\mu} = \mathbf{q}$. In the more general case when the views are not certain, the posterior estimate of $\boldsymbol{\mu}$ given the prior $\boldsymbol{\mu} \sim N(\boldsymbol{\pi}, \mathbf{Q})$ and the views $\mathbf{P}\boldsymbol{\mu} = \mathbf{q} + \boldsymbol{\epsilon}$, with $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Omega})$, is

$$\hat{\boldsymbol{\mu}} = \boldsymbol{\pi} + \mathbf{Q}\mathbf{P}^\top \left(\mathbf{P}\mathbf{Q}\mathbf{P}^\top + \boldsymbol{\Omega}\right)^{-1} (\mathbf{q} - \mathbf{P}\boldsymbol{\pi}). \tag{7.3}$$

When $\boldsymbol{\Omega}$ is non-singular, $\hat{\boldsymbol{\mu}}$ also has the following equivalent expression:

$$\hat{\boldsymbol{\mu}} = \left(\mathbf{Q}^{-1} + \mathbf{P}^\top \boldsymbol{\Omega}^{-1}\mathbf{P}\right)^{-1} \left(\mathbf{Q}^{-1}\boldsymbol{\pi} + \mathbf{P}^\top \boldsymbol{\Omega}^{-1}\mathbf{q}\right) \tag{7.4}$$

Observe that the expression (7.2) for certain views can be recovered from (7.3) by taking $\boldsymbol{\Omega} = \mathbf{0}$.

We next give a derivation of the formula (7.4) for the posterior estimate of $\boldsymbol{\mu}$ when $\boldsymbol{\Omega}$ is non-singular. The exercises at the end of the chapter show how the derivation can be tweaked for any $\boldsymbol{\Omega}$. Stack the two equations for the market equilibrium $\boldsymbol{\pi} = \boldsymbol{\mu} + \boldsymbol{\epsilon}_{\boldsymbol{\pi}}, \boldsymbol{\epsilon}_{\boldsymbol{\pi}} \sim N(\mathbf{0}, \mathbf{Q})$, and for the investor's views $\mathbf{q} = \mathbf{P}\boldsymbol{\mu} + \boldsymbol{\epsilon}_{\mathbf{q}}$, $\boldsymbol{\epsilon}_{\mathbf{q}} \sim N(\mathbf{0}, \boldsymbol{\Omega})$ as

$$\mathbf{y} = \mathbf{M}\boldsymbol{\mu} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$$

for

$$\mathbf{y} = \begin{bmatrix} \boldsymbol{\pi} \\ \mathbf{q} \end{bmatrix}, \quad \mathbf{M} = \begin{bmatrix} \mathbf{I} \\ \mathbf{P} \end{bmatrix}, \quad = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \end{bmatrix}.$$

The estimation problem can be stated as the following weighted least-squares problem:

$$\min_{\boldsymbol{\mu}} (\mathbf{y} - \mathbf{M}\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{M}\boldsymbol{\mu}).$$

The optimality conditions for this problem yield

$$2\mathbf{M}^\top \boldsymbol{\Sigma}^{-1} \mathbf{M}\boldsymbol{\mu} - 2\mathbf{M}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} = \mathbf{0}$$

Hence we obtain

$$\begin{aligned}
\hat{\mu} &= \left(\mathbf{M}^\top \boldsymbol{\Sigma}^{-1} \mathbf{M}\right)^{-1} \mathbf{M}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} \\
&= \left(\mathbf{Q}^{-1} + \mathbf{P}^\top \boldsymbol{\Omega}^{-1} \mathbf{P}\right)^{-1} \left(\mathbf{Q}^{-1} \boldsymbol{\pi} + \mathbf{P}^\top \boldsymbol{\Omega}^{-1} \mathbf{q}\right) \\
&= \boldsymbol{\pi} + \mathbf{Q}\mathbf{P}^\top \left(\mathbf{P}\mathbf{Q}\mathbf{P}^\top + \boldsymbol{\Omega}\right)^{-1} (\mathbf{q} - \mathbf{P}\pi)
\end{aligned}$$

General Remarks

First we note that the Black-Litterman model can be thought of as an "inverse optimization problem":

If one views the market equilibrium as the optimum solution of a portfolio optimization problem, what data would produce this outcome? In particular, given investor views, what choice of $\boldsymbol{\mu}$ would best fit the market equilibrium solution? This leads to the least-squares problem formulated above, whose solution $\hat{\boldsymbol{\mu}}$ has the expression (7.3) that we just computed. This "inverse optimization" philosophy was proposed by Bertsimas et al. (2012). It has the added flexibility of allowing investor views on volatility and market dynamics.

We also note that the analysis in the Black-Litterman model relies heavily on the assumption that the error term $\epsilon \sim N(\mathbf{0}, \mathbf{Q})$ is normally distributed. When this is not the case, the Black-Litterman framework is still meaningful but the analysis is more complex and a closed-form solution like (7.3) does not usually exist.

Shrinkage Estimation

Another approach to improve the quality of estimated expected returns is based on shrinkage estimators. These types of estimators are rooted in the classical finding of Stein (1956) that biased estimators may, in a formal fashion, be superior to the unbiased sample mean. As we detail below, the central idea is that the estimation can be improved by shrinking the sample mean towards a target. The non-technical article of Efron and Morris (1977) gives an enlightening discussion of an application of this approach to the estimation of baseball batting averages.

To formalize ideas, consider the problem of estimating the mean of an N dimensional multivariate normal variable $\mathbf{r} \sim N(\boldsymbol{\mu}, \mathbf{V})$ from a set of observations $\mathbf{r}_1, \ldots, \mathbf{r}_T$. For a given estimate $\hat{\boldsymbol{\mu}}$, consider the quadratic loss function

$$L(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}) := (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\top \mathbf{V}^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}). \tag{7.5}$$

For a given loss function, the risk of an estimator is $\mathbb{E}(L(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}))$, where the expectation is taken over the space of samples $\mathbf{r}_1, \ldots, \mathbf{r}_T$. An estimator is inadmissible if there exists another estimator with lower risk.

For the quadratic loss function (7.5) and $N = 1, 2$, it is known that the optimal estimator is the sample mean $\bar{\mathbf{r}} := (1/T)(\mathbf{r}_1 + \cdots + \mathbf{r}_T)$. By contrast, for $N > 2$, the James-Stein shrinkage estimator

$$\hat{\boldsymbol{\mu}}_{JS} := (1 - w)\bar{\mathbf{r}} + w\mu_0 \mathbf{1}$$

has lower risk than the sample mean $\bar{\mathbf{r}}$ for

$$w = \min\left(1, \frac{N-2}{T(\bar{\mathbf{r}} - \mu_0 \mathbf{1})^\top \mathbf{V} (\bar{\mathbf{r}} - \mu_0 \mathbf{1})}\right).$$

Here T is the number of observations, and μ_0 is an arbitrary number. The vector $\mu_0 \mathbf{1}$ and the weight w are referred to as the shrinkage target and shrinkage factor respectively. Although some choices of μ_0 are better than others, what is surprising is that in theory μ_0 could be any fixed number. This fact is called the Stein paradox.

The James-Stein shrinkage estimator can be seen as a combination of two estimators:

- an estimator with little or no structure (like the sample mean);
- an estimator with a lot of structure (the shrinkage target).

The exact combination of these two estimators is determined by a certain shrinkage intensity. As we discuss below, this same shrinkage approach has been successfully applied to obtain improved estimators of covariance matrices and beta exposures.

The following shrinkage estimator proposed by Jorion (1986) is fairly popular in the financial literature. The estimator was derived via an empirical Bayesian approach. As a shrinkage target, use the vector $\mu_0 \mathbf{1}$ for

$$\mu_0 := \frac{\bar{\mathbf{r}}^\top \mathbf{V}^{-1} \mathbf{1}}{\mathbf{1}^\top \mathbf{V}^{-1} \mathbf{1}},$$

and as a shrinkage intensity, use

$$w := \frac{N+2}{N+2+T\left(\bar{\mathbf{r}}-\mu_0 \mathbf{1}\right)^\top \mathbf{V}^{-1}\left(\bar{\mathbf{r}}-\mu_0 \mathbf{1}\right)}$$

Shrinkage can also be applied to other estimation problems. For instance, Ledoit and Wolf (2003, 2004) propose shrinkage approaches for covariance estimation in the same spirit as the James-Stein shrinkage estimator: shrink the sample covariance matrix $\overline{\mathbf{V}}$ (an unstructured estimator) towards a highly structured target estimator $\mathbf{V}_0$:

$$\hat{\mathbf{V}}_{LW} := (1 - w)\overline{\mathbf{V}} + w\mathbf{V}_0.$$

The shrinkage target estimator $\mathbf{V}_0$ could be a single-factor estimator, an estimator of the covariance matrix with constant correlation, a diagonal matrix, or a multiple of the identity matrix.

Shrinkage is also routinely used for estimating benchmark exposures in a stock universe. Let $r_i, i = 1, \ldots, N$, denote the excess returns of stocks in the investment universe and r_B denote the excess return of the benchmark. Recall that the beta of stock i captures the benchmark portion of the return on stock i via the linear model

$$r_i = \beta_i r_B + \theta_i.$$

The value of β_i is the benchmark exposure of stock i. Given historical realizations of r_i, for $i = 1, \ldots, N$, and r_B, we can obtain estimates $\hat{\beta}_i$ of β_i for $i = 1, \ldots, N$ via ordinary least-squares linear regression. The forecasts given by these natural estimators tend to overestimate the betas of stocks with high benchmark exposure and underestimate the betas of the stocks with low benchmark exposure. Improved forecasts can be obtained by shrinking the betas obtained from the least-squares procedures towards one (the benchmark beta):

$$\hat{\boldsymbol{\beta}} = (1 - w)\overline{\boldsymbol{\beta}} + w\mathbf{1},$$

where $\overline{\boldsymbol{\beta}}$ denotes the vector of beta estimates from the least-squares procedure.

A common rule of thumb in the above shrinkage estimators of covariance matrix and vector of betas is to use a shrinkage intensity $w = 1/2$ for estimates based on 60 -month long historical data. A thorough discussion on the appropriate choice of shrinkage intensity can be found in Ledoit and Wolf (2003, 2004) and Blume (1975). Portfolio optimization can be viewed as a stochastic optimization problem (see Chapter 10). Shrinkage is relevant in this more general context as well (Davarnia and Cornuéjols, 2017).

Resampled Efficiency

A different approach to address the sensitivity of mean-variance optimization to estimation error is to apply the bootstrap technique from statistics. The bootstrap technique is a method to estimate standard errors and confidence intervals of statistics of a dataset via random resampling from the dataset with replacement.

The application of bootstrapping to mean-variance optimization was initially explored by Jorion (1992) and later further developed and marketed by Michaud and Michaud (2008). The basic idea is to consider the joint problem of parameter estimation and portfolio construction as a statistical procedure: the efficient portfolios can be seen as a statistic on a set of financial data used for estimation. The resampled efficiency technique proposed by Michaud and Michaud proceeds by applying bootstrapping to this statistical process. Suppose there is a procedure that estimates the vector of expected returns and covariance from historical data. Use the available data to produce these estimates and compute efficient portfolios. Repeat this same process either by sampling from these estimates, or by bootstrapping the available data to obtain new estimates of expected returns and covariances.

All these estimates are statistically equivalent. For each of them, we can generate the corresponding set of efficient portfolios. The collection of all of these portfolios forms some sort of equivalence region. We would like to take some average of the equivalence region so that the effects of estimation error are mitigated. However, it is not obvious how to average since the equivalence region contains portfolios with low and high variance. We do not want to mix "apples and oranges". Michaud and Michaud's suggestion is to average portfolios that are in some equivalent risk-return bucket. To that end, we propose the following procedure: For each efficient frontier, save m evenly distributed efficient portfolios. Rank them 1 to m. Then take averages of same-rank portfolios from all efficient frontiers.

This resampling procedure can be more precisely described as follows (see Algorithm 7.1). Suppose we have a procedure to produce estimates $\hat{\mu}$ and $\hat{\mathbf{V}}$ from a history of T periods of historical data.

Algorithm 7.1 Resampling procedure

1. **for** $i = 1, \ldots, S$ **do**
2. simulate a new history of T periods by resampling the original history
3. use the simulated history to generate new estimates $\hat{\mu}_i$ and $\hat{V}_i$
4. use $\hat{\mu}_i$ and $\hat{V}_i$ to generate m equally spaced efficient portfolios $\mathbf{x}_{1,i}, \ldots, \mathbf{x}_{m,i}$
5. **end for**

To generate the resampled efficient portfolios, take averages of equally ranked efficient portfolios generated above:

$$\mathbf{x}_{j,\text{ resampled}} := \frac{1}{S}\sum_{i=1}^{S}\mathbf{x}_{j,i}.$$

The resampled efficient frontier is the expected return versus standard deviation chart of the resampled efficient portfolios with the original estimates $\hat{\mu}$ and $\hat{\mathbf{V}}$.

There are a number of limitations to resampling (Scherer, 2002). The entire process is only a heuristic; there is no sound theory to support why the process should mitigate the effects of estimation error. The methodology does have the feature of generating portfolios that look well diversified, and this is generally well received. However, this feature can be attributed to the role of variability in the averaging process. The process is intense computationally, as it multiplies the work involved in conventional mean-variance optimization. Furthermore, the procedure does not provide any clear mechanism to facilitate the incorporation of views as in the Black-Litterman model.

Robust Optimization

Robust optimization is a fairly recent development that considers uncertainty in some parameters directly in the optimization problem. The general idea of robust optimization is to generate a solution that is good for all possible realizations of the uncertain parameters. Consider a minimization problem with inequality constraints

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}, \mathbf{p}) \\ \text{s.t.} & g_i(\mathbf{x}, \mathbf{p}) \leq 0, i = 1, \ldots, m \end{array} \tag{7.6}$$

Here the vector **p** stands for some parameters that define the objective and constraints functions.

Consider first the case when the uncertain parameters occur in the constraints only. Assume that the set of parameters **p** is uncertain but it is known to be in some uncertainty set $\mathcal{U}$. In this case a robust version of (7.6) is one where the optimization is performed over points that are feasible for all possible realizations of the uncertain parameters $\mathbf{p} \in \mathcal{U}$; that is,

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{s.t.} & \max_{\mathbf{p} \in \mathcal{U}} g_i(\mathbf{x}, \mathbf{p}) \leq 0, i = 1, \ldots, m \end{array}$$

On the other hand, consider the case when the uncertain parameters occur in the objective only. In this case a robust version of (7.6) is one that finds the solution that would be best, given the worst possible realization of the uncertain parameters $\mathbf{p} \in \mathcal{U}$; that is,

$$\begin{array}{rl} \min_{\mathbf{x}} \max_{\mathbf{p} \in \mathcal{U}} & f(\mathbf{x}, \mathbf{p}) \\ \text{s.t.} & g_i(\mathbf{x}) \leq 0, i = 1, \ldots, m \end{array}$$

If uncertain parameters occur in both the objective and constraints, then the robust version is as follows:

$$\begin{array}{rl} \min_{\mathbf{x}} \max_{\mathbf{p} \in \mathcal{U}} & f(\mathbf{x}, \mathbf{p}) \\ \text{s.t.} & \max_{\mathbf{p} \in \mathcal{U}} g_i(\mathbf{x}, \mathbf{p}) \leq 0, i = 1, \ldots, m. \end{array}$$

As we detail in Chapter 19, for suitable types of uncertainty sets the above robust versions can be rewritten as an optimization problem that is manageable albeit via more involved optimization machinery.

Other Diversification Approaches

The challenges associated with expected return estimation and the input sensitivity of mean-variance models have given rise to quantitative portfolio construction approaches that eschew expected return estimation and focus on managing risk only. We next discuss some popular approaches of this kind that have led to the development of a variety of investment products in the asset management industry.

Assume $\mathbf{V}$ is the covariance matrix of asset returns in some investment universe and $\boldsymbol{\sigma}$ is the vector of volatilities (standard deviations) of the asset returns. In particular, the diagonal entries of $\mathbf{V}$ are the squares of the entries of $\boldsymbol{\sigma}$.

The minimum-risk portfolio is the portfolio in the efficient frontier of minimum variance. In the absence of constraints, this portfolio is the solution to the following quadratic programming model:

$$\begin{array}{ll} \min_{\mathbf{x}} & \mathbf{x}^\top \mathbf{V} \mathbf{x} \\ \text{s.t.} & \mathbf{1}^\top \mathbf{x} = 1. \end{array}$$

That is,

$$\mathbf{x}^* = \frac{1}{\mathbf{1}^\top \mathbf{V}^{-1}\mathbf{1}} \mathbf{V}^{-1}\mathbf{1}$$

For the special case $\mathbf{V} = \mathbf{I}$ (the $N \times N$ identity matrix) the minimum-risk portfolio is the so-called equally weighted portfolio

$$x_i^* = \frac{1}{N}, \quad i = 1, \ldots, N$$

where N is the number of assets in the universe. On the other hand, if $\mathbf{V}$ is diagonal, that is, $\mathbf{V} = \text{diag}(\boldsymbol{\sigma})^2$, then the portfolio components are proportional to the inverse of the squares of the volatilities:

$$x_i^* = \frac{1/\sigma_i^2}{\sum_{i=1}^N 1/\sigma_i^2}, \quad i = 1, \ldots, N$$

In particular, this portfolio is the value-weighted portfolio if the capitalization of asset i is used as a proxy for $1/\sigma_i^2$.

Risk Parity

We next discuss two more recent diversification approaches, namely risk parity and maximum diversification. To that end, we first discuss the related concept of risk contribution. Observe that the risk (standard deviation) of a portfolio $\mathbf{x} = (x_1, \ldots, x_N)$ is given by

$$\sigma_P(\mathbf{x}) = \sqrt{\mathbf{x}^\top \mathbf{V}\mathbf{x}}$$

If we compute the partial derivative of this portfolio with respect to x_i, we obtain the marginal contribution to risk of asset i :

$$MCR_i(\mathbf{x}) = \frac{\partial \sigma_P(\mathbf{x})}{\partial x_i} = \frac{(\mathbf{V}\mathbf{x})_i}{\sqrt{\mathbf{x}^\top \mathbf{V}\mathbf{x}}}, \quad i = 1, \ldots, N$$

The contribution to risk of asset i is:

$$CR_i(\mathbf{x}) = x_i \cdot MCR_i(\mathbf{x}) = \frac{x_i \cdot (\mathbf{V}\mathbf{x})_i}{\sqrt{\mathbf{x}^\top \mathbf{V}\mathbf{x}}}, \quad i = 1, \ldots, N$$

Observe that

$$\sum_{i=1}^{N} CR_i(\mathbf{x}) = \sqrt{\mathbf{x}^\top \mathbf{V}\mathbf{x}} = \sigma_P(\mathbf{x})$$

Risk Parity

Consequently, we say that $\mathbf{x}$ is a risk-parity portfolio if all the assets in the portfolio have the same contribution to risk; that is, if

$$CR_i(\mathrm{x}) = \frac{\sigma_P(\mathrm{x})}{N}, \quad i = 1, \ldots, N$$

Again, in the special case when $\mathbf{V} = \mathrm{diag}(\boldsymbol{\sigma})^2$, the fully invested risk-parity portfolio is

$$x_i^* = \frac{1/\sigma_i}{\sum_{i=1}^N 1/\sigma_i}, \quad i = 1, \ldots, N.$$

Risk Parity

For a general covariance matrix $\mathbf{V}$ and portfolio constraints, it may not be possible to attain perfect risk parity. In this case, we can instead minimize some kind of measure of deviation from risk parity. Here are some choices for examples of these kinds of measures proposed in the literature:

$$DRP_1(\mathbf{x}) = \sum_{i=1}^{N} \sum_{j=1}^{N} \left(x_i \cdot (\mathbf{V}\mathbf{x})_i - x_j \cdot (\mathbf{V}\mathbf{x})_j\right)^2$$

$$DRP_2(\mathbf{x}) = \sum_{i=1}^{N} \left(\frac{x_i \cdot (\mathbf{V}\mathbf{x})_i}{\mathbf{x}^\top \mathbf{V}\mathbf{x}} - \frac{1}{N}\right)^2$$

$$DRP_3(\mathbf{x}) = \sum_{i=1}^{N} \left|\frac{x_i \cdot (\mathbf{V}\mathbf{x})_i}{\mathbf{x}^\top \mathbf{V}\mathbf{x}} - \frac{1}{N}\right|$$

The optimization problem associated with minimizing any of these deviation measures is in general quite a bit more challenging than other mean-variance models, as these problems are not convex. The development of efficient numerical algorithms to solve these kinds of optimization problems is a topic of current research.

Maximum Diversification

Another approach to diversification is maximum diversification (Choueifaty and Coignard, 2008). More precisely, maximize the diversification ratio

$$\frac{\sigma^\top \mathbf{x}}{\sqrt{\mathbf{x}^\top \mathbf{V} \mathbf{x}}}$$

where $\boldsymbol{\sigma}$ is the vector of asset volatilities. A motivation for this approach can be given as follows: Observe that the diversification ratio is proportional to the Sharpe ratio if $\boldsymbol{\mu}$ is proportional to $\boldsymbol{\sigma}$. Hence maximizing diversification is equivalent to maximizing the Sharpe ratio under the assumption that the expected returns of the assets are proportional to their volatilities.

Maximum Diversification

In the absence of other constraints, the fully invested maximum diversification portfolio is the solution to the optimization problem

$$\begin{array}{ll} \min_{\mathbf{x}} & \frac{\boldsymbol{\sigma}^\top \mathbf{x}}{\sqrt{\mathbf{x}^\top \mathbf{x}}} \\ \text{s.t.} & \mathbf{1}^\top \mathbf{x} = 1 \end{array} \tag{7.7}$$

In the special case when $\mathbf{V} = \text{diag}(\boldsymbol{\sigma})^2$, the solution to (7.7) coincides with the fully invested risk-parity portfolio:

$$x_i^* = \frac{1/\sigma_i}{\sum_{i=1}^N 1/\sigma_i}, \quad i = 1, \ldots, N$$