FinMath Topic 2

Feasible Set of a Market and Markowitz Mean-variance Theory

Updated on August 21, 2022

© 2022 Nicky老师
nickycanhelp@qq.com

In this topic, we first demonstrate how to construct the feasible set of multiple risky assets from a practical view and then revisit this problem by introducing the seminal Markowitz mean-variance model for readers who are willing to go deeper.

I. Preliminaries

1 Estimations on mean and covariance of returns

Consider a general market with n risky assets, $n = 2, 3, \ldots$, and their returns are denoted by a joint random vector $R = (R_1, R_2, \ldots, R_n)' \in \mathbb{R}^n$. Theoretically, the mean vector $\mu \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ of R are defined, respectively, as

$$\mu = \mathbb{E}[R] = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}, \ \Sigma = \mathrm{Cov}(R) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix},$$

where $\mu_i = \mathbb{E}[R_i]$ and $\sigma_{ij} = \mathrm{Cov}(R_i, R_j)$ for $i, j = 1, 2, \ldots, n$, and σ_{ij} could be further written as $\rho_{ij}\sigma_i\sigma_j$ where ρ_{ij} is the correlation between returns of two assets i and j and σ_i is the volatility of asset i. We know that the covariance matrix Σ is symmetric[1], and we further assume that Σ is invertible.

◊ 注意, 本文遵循惯例将向量当做列向量处理 (除非特别说明是行向量), 本文转置符号写作 $'$ (而其他常见符号如 ^{T}), 因而本文求导符号一般会写作$df(x)/dx$而非$f'(x)$. 当(列)向量出现在段落文字内, 其会以行向量形式出现但在结尾带有转置符号$'$以强调其列向量本质, 此举纯粹是为了节省空间 (试想一下非要在段落内写成列的形式...)

In practice, we often estimate μ and Σ from the historical data. More precisely, suppose we have already collected a dataset $X \in \mathbb{R}^{m\times n}$,

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix},$$

[1] Its diagonals are just variances of each return while the off-diagonals are covariance between two distinct returns which satisfies $\sigma_{ij} = \sigma_{ji}$.

whose columns are the time series of historical returns, namely, $(x_{1j}, x_{2j}, \ldots, x_{mj})'$ represents the realized returns of asset j in the past m trading days. Then μ could be estimated from the sample mean $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_n)'$, where

$$\hat{\mu}_i = \frac{1}{m}\sum_{t=1}^{m} x_{ti}, \ i = 1, 2, \ldots, n, \tag{1}$$

and Σ could be estimated from the sample covariance $\hat{\Sigma}$ whose ith-row and jth-column element is given by

$$\hat{\sigma}_{ij} = \frac{1}{m-1}\sum_{t=1}^{m}(x_{ti} - \hat{\mu}_i)(x_{tj} - \hat{\mu}_j), \ i, j = 1, 2, \ldots, n. \tag{2}$$

◊ 在求sample covariance时除以$m-1$而非m的原因: 1)因公式中用到了求过的sample mean, 导致该系列数据的自由度(degree of freedom)减一(可以理解成: 在给定sample mean的时候, 只有$m-1$个数据点是可以任取的,"自由"的); 2)除以$m-1$才使得该表达式是covariance的unbiased estimator(统计学知识).

In fact, there is a smart and compact way to compute $\hat{\mu}$ and $\hat{\Sigma}$ by using standard matrix multiplications. To see this, let us denote by $\mathbb{1}_m \in \mathbb{R}^m$ the column vector whose elements are all ones. Then we have

$$\hat{\mu} = \left(\frac{1}{m}\mathbb{1}_m' X\right)' = \frac{1}{m}X'\mathbb{1}_m, \tag{3}$$

$$\hat{\Sigma} = \frac{1}{m-1}(X'X - m\hat{\mu}\hat{\mu}'). \tag{4}$$

⚠ 上述两公式正确的前提是, 数据集X必须按照文中所述的方式排列, 也即每个资产的收益率时间序列数据作为X的一列出现.

Three tips in regard to the important formulas (3) and (4):

- If you are not comfortable with these two compact forms of calculations, then just compute each element of $\hat{\mu}$ and $\hat{\Sigma}$ one by one through (1) and (2) (this will be a tedious and time-consuming job);
- and if you are fine with the expressions but do not want to know (or has been familiar with) the details behind them, then just take them for granted and go for coding them;
- but if you are interested in their derivations, then let us verify them in the following.

We write the compact forms in more details, namely, by matrix multiplications we have

$$\mathbb{1}_m' X = (1, 1, \ldots, 1)\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} = \left(\sum_{t=1}^{m} x_{t1}, \sum_{t=1}^{m} x_{t2}, \ldots, \sum_{t=1}^{m} x_{tn}\right),$$

and this explains how we summarize from (1) to (3). Besides, according to the definition of covariance in multiple dimension, i.e.,

$$\Sigma = \mathbb{E}[(R - \mathbb{E}[R])(R - \mathbb{E}[R])'], \ R \in \mathbb{R}^n,$$

we have

$$\hat{\Sigma} = \frac{1}{m-1}\sum_{t=1}^{m}(r_t - \hat{\mu})(r_t - \hat{\mu})', \tag{5}$$

where r_t is the tth-day observed returns of n assets (in column form), namely,

$$r_t = (x_{t1}, x_{t2}, \ldots, x_{tn})',\ t = 1, 2, \ldots, m.$$

◊ 此处, 把X的第t行写成列向量的形式就是r_t.

Therefore, we have some useful results

$$(r_1, r_2, \ldots, r_m) = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{m1} \\ x_{12} & x_{22} & \cdots & x_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{mn} \end{pmatrix} = X',$$

$$\sum_{t=1}^{m} r_t r_t' = (r_1, r_2, \ldots, r_m)\begin{pmatrix} r_1' \\ r_2' \\ \vdots \\ r_m' \end{pmatrix} = X'X,$$

$$\sum_{t=1}^{m} r_t = (r_1, r_2, \ldots, r_m)\mathbb{1}_m = X'\mathbb{1}_m = m\hat{\mu}.$$

◊ 此处用到一些block matrix的处理方法.

Then we continue from (5) and obtain

$$\begin{aligned}\hat{\Sigma} &= \frac{1}{m-1}\sum_{t=1}^{m}(r_t r_t' - r_t\hat{\mu}' - \hat{\mu}r_t' + \hat{\mu}\hat{\mu}') \\ &= \frac{1}{m-1}\left(\sum_{t=1}^{m} r_t r_t' - (\sum_{t=1}^{m} r_t)\hat{\mu}' - \hat{\mu}(\sum_{t=1}^{m} r_t)' + m\hat{\mu}\hat{\mu}'\right) \\ &= \frac{1}{m-1}\left(\sum_{t=1}^{m} r_t r_t' - m\hat{\mu}\hat{\mu}' - m\hat{\mu}\hat{\mu}' + m\hat{\mu}\hat{\mu}'\right) \\ &= \frac{1}{m-1}(X'X - m\hat{\mu}\hat{\mu}').\end{aligned}$$

2 Statistics of portfolio return

We review the compact formulas to calculate the mean and variance of the random return of a portfolio which is formed by multiple assets. Suppose we are provided the same market settings as in Section 1, and we construct a portfolio by some weights on these assets, denoted by $w = (w_1, \ldots, w_n)'$. Then the portfolio return R_P can be written as

$$R_P = w_1 R_1 + w_2 R_2 + \cdots + w_n R_n = w'R.$$

Therefore, the expected portfolio return and the variance of R_P is computed, respectively, through

$$\mathbb{E}[R_P] = \mathbb{E}[w'R] = w'\mu, \tag{6}$$
$$\mathrm{Var}(R_P) = \mathrm{Var}(w'R) = w'\Sigma w. \tag{7}$$

It is easy to understand (6) and we would like to explain (7) a little bit from two aspects, if you are not familiar with this expression:

- We write variance by covariance and get

$$\begin{aligned}
\mathrm{Var}(w'R) &= \mathrm{Cov}\left(\sum_{i=1}^n w_i R_i, \sum_{i=1}^n w_i R_i\right) = \sum_{i,j=1}^n w_i w_j \mathrm{Cov}(R_i, R_j) \\
&= (w_1, w_2, \ldots, w_n) \begin{pmatrix} \mathrm{Cov}(R_1,R_1) & \mathrm{Cov}(R_1,R_2) & \cdots & \mathrm{Cov}(R_1,R_n) \\ \mathrm{Cov}(R_2,R_1) & \mathrm{Cov}(R_2,R_2) & \cdots & \mathrm{Cov}(R_2,R_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(R_n,R_1) & \mathrm{Cov}(R_n,R_2) & \cdots & \mathrm{Cov}(R_n,R_n) \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} \\
&= w'\Sigma w.
\end{aligned}$$

- For a general random vector $AX + b$ where $X \in \mathbb{R}^n$ is random and $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ are constant matrices, we have

$$\begin{aligned}
&\mathrm{Cov}(AX+b) = \mathbb{E}[(AX+b)(AX+b)'] - \mathbb{E}[(AX+b)]\mathbb{E}[(AX+b)'] \\
&= \mathbb{E}[AXX'A' + AXb' + bX'A' + bb'] - (A\mathbb{E}[X] + b)(\mathbb{E}[X']A' + b') \\
&= A\mathbb{E}[XX']A' + A\mathbb{E}[X]b' + b\mathbb{E}[X']A' + bb' \\
&\quad - A\mathbb{E}[X]\mathbb{E}[X']A' - A\mathbb{E}[X]b' - b\mathbb{E}[X']A' - bb' \\
&= A(\mathbb{E}[XX'] - \mathbb{E}[X]\mathbb{E}[X'])A' \\
&= A\mathrm{Cov}(X)A'.
\end{aligned}$$

  Then we derive $\mathrm{Cov}(w'R) = w'\mathrm{Cov}(R)w = w'\Sigma w$.

◊ 上述variance计算公式的助记方法: 组合收益率R_P是一维随机变量, 所以其variance一定是一个scalar, 这和矩阵运算$w'\Sigma w$的scalar结果一致. 该公式告诉我们: 组合收益率R_P的方差是关于weight w的quadratic function.

3 Lagrangian method

From elementary math course we know that, if we would like to find the extrema (maxima or minima) of a function, we often set the first-order derivative to be zero as the necessary condition of optimality. More specifically, for an *unconstrained* optimization problem,

- if we want to minimize a single-variable function $f(x)$, then it is necessary that

$$\left.\frac{df(x)}{dx}\right|_{x=x^*} = 0$$

  for x^* being a minimum point of $f(x)$;

- and similarly, if we want to minimize a multi-variable function $f(x_1, x_2, \ldots, x_n)$, then at a minimum point of the function, it is necessary that

$$\frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_i} = 0,\ i = 1, 2, \ldots, n.$$

As for an *constrained* optimization problem with *equality* constraints, namely,

$$\begin{array}{rl} \text{minimize} & f(x_1, x_2, \ldots, x_n) \\ \text{subject to} & g_k(x_1, x_2, \ldots, x_n) = 0,\ k = 1, 2, \ldots, K, \end{array}$$

a well-known strategy for dealing with this type of problems is known as the Lagrangian method. The basic idea is to convert the original constrained problem into an unconstrained counterpart such that the previous derivative test could be applied. More precisely, we first introduce K *Lagrange multipliers*, $\lambda_1, \lambda_2, \ldots, \lambda_K$, for each equality constraint and form the so-called *Lagrangian function* as below,

$$L(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_K) = f(x_1, x_2, \ldots, x_n) - \sum_{k=1}^{K} \lambda_k g_k(x_1, x_2, \ldots, x_n).$$

Then we directly minimize the Lagrangian function $L(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_K)$ w.r.t. $n + K$ variables just as what we do for an unconstrained problem, and we would obtain a system of equations given by

$$\begin{aligned} \frac{\partial L}{\partial x_i} &= 0,\ i = 1, \ldots, n, \\ \frac{\partial L}{\partial \lambda_k} &= 0,\ k = 1, \ldots, K, \end{aligned}$$

and solving the above $n + K$ equations for $n + K$ unknowns, we would get the necessary conditions for the original n unknowns satisfying the equality constraints as well. We ignore the theoretical evidence behind this method[2] but just emphasize on how to use it later in this topic.

[2] In fact, the Lagrangian method is a special case of the famous Karush–Kuhn–Tucker conditions, which plays a very important role in optimization field.

II. Feasible Region: From Three Assets to Multiple

After we study the simple two-asset case in Topic1, we now move forward to talk about the feasible region constructed by three assets and more. A comprehensive investigation on the feasible region of components from a stock index will be given through Matlab. We will also build up the efficient frontier in the mean-variance sense by using the real market data.

4 Feasible region among three risky assets

Recall that in Topic1 when we want to figure out the feasible set among two risky and one risk-free assets, we first construct the feasible set between two risky assets as a parabolic curve, and then connect a (risky) portfolio on the curve with the riskless asset by straight line, and finally realize that by varying this portfolio along the curve, we could trace out the feasible region as desired (see Figure 7 in Topc1). This smart way can be duplicated here when we discuss about the feasible set among three risky assets. Let us explain the procedure by concrete examples.

For simplicity, we first focus on the market condition where shorting is not allowed. Suppose we are given three (artificial) risky assets A, B and C with means and standard deviations of and correlations between returns R_A, R_B, R_C given by

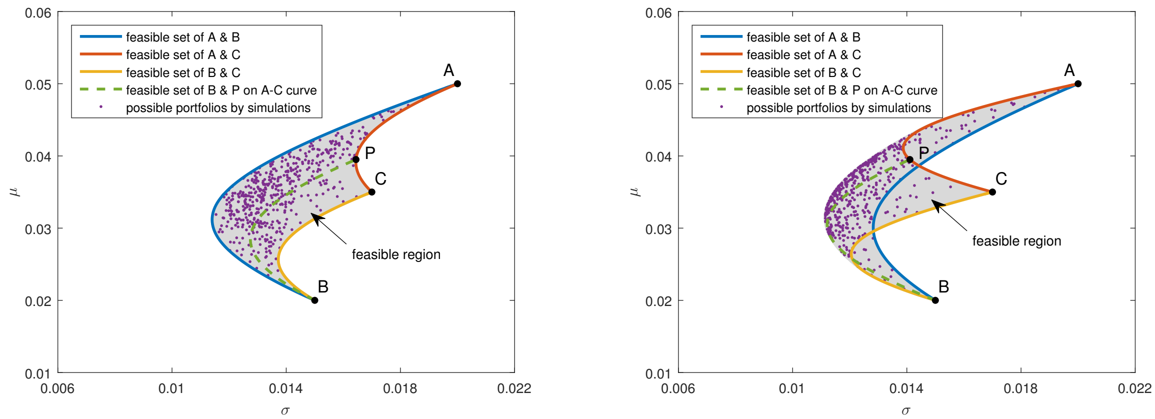
$$\begin{aligned} \mu_A &= 0.05,\ \sigma_A = 0.02, \\ \mu_B &= 0.02,\ \sigma_B = 0.015, \\ \mu_B &= 0.035,\ \sigma_B = 0.017, \\ \rho_{AB} &= -0.1,\ \rho_{AC} = 0.65,\ \rho_{BC} = 0.5. \end{aligned}$$

It is easy to draw the pairwise feasible curves among three assets, as shown in Figure 1(a). Note that since we confine ourselves in the world of no shoring, all the curves cannot be extended beyond the three terminal points. We now pick up a two-component portfolio P on the A-C curve and consider how to form a three-component portfolio when combining with B.[3] The result turns out to be another parabolic curve, which just lies in between A-B curve and C-B curve. And if we move P along the A-C curve and link with B every time, we could trace out a feasible region of three risky assets A, B, C as shaded in gray in the figure, which is enclosed by the three curves. Moreover, we randomly construct some portfolios based on these three assets to check whether they lie in the region. This is done by simulating the weights w_A, w_B, w_C according to the constraints[4]

$$\begin{aligned} w_A + w_B + w_C &= 1, \\ w_A, w_B, w_C &\geq 0, \end{aligned}$$

[3] A normal way to do this is that we first compute $\mu_P, \sigma_P, \rho_{PB}$ numerically and then return to the standard procedure for two risky assets.

[4] The method of simulation will be explained in the next section through Matlab.



(a) Example I (b) Example II

Figure 1: Two typical shapes of feasible region for three risky assets when shorting is forbidden (shaded in gray).

and then computing the expected return and the volatility of the resulted portfolio (denoted by Q, for example) through

$$\begin{aligned}\mu_Q &= \mathbb{E}[R_Q] = \mathbb{E}[w_A R_A + w_B R_B + w_C R_C] \\ &= w_A\mu_A + w_B\mu_B + w_C\mu_C, \\ \sigma_Q^2 &= \mathrm{Var}(R_Q) = \mathrm{Var}(w_A R_A + w_B R_B + w_C R_C) \\ &= \sum_{i,j\in\{A,B,C\}} w_i w_j \sigma_{ij}, \text{ where } \sigma_{ij} = \rho_{ij}\sigma_i\sigma_j.\end{aligned}$$

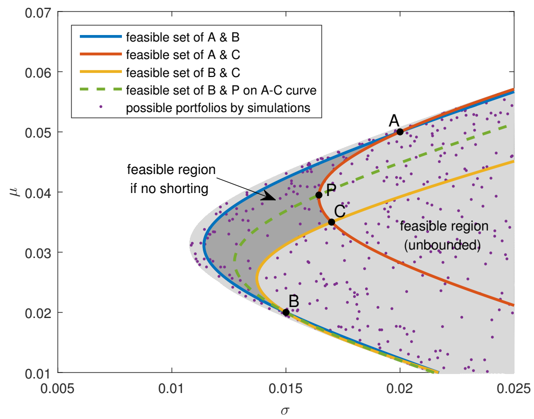
After calculations, all the points (σ_Q, μ_Q) from simulations (colored in purple) undoubtedly scatter in the shaded region of Figure 1(a).

In the above example, we see a typical (but not unique) layout of the pairwise curves where there is one curve dominating the rest and no intersection among them. Definitely, there are other types of layout, depending on the statistics of the given assets. For instance, let us consider another example with the same mean and volatility as above but different correlations which are given by

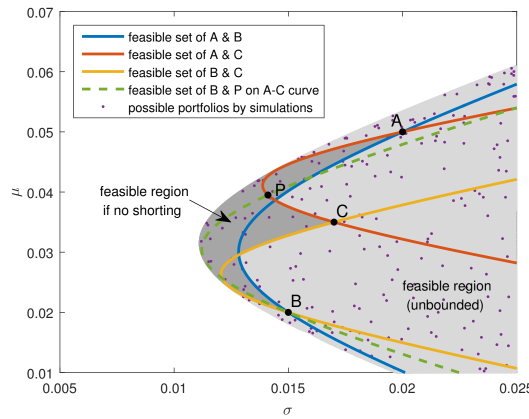
$$\rho_{AB} = \rho_{AC} = \rho_{BC} = 0.15.$$

We then calculate and depict the pairwise feasible curves on the σ-μ graph, as shown in Figure 1(b), and observe that in this case no one can continuously dominate the others and thus they share some intersections. We follow the similar procedure to find the feasible region among these new risky assets and figure out that this time the shaded region still has a parabolic frontier from the left which however is not anyone of the original pairwise curves. We also simulate plenty of possible portfolios just as previous case in order to demonstrate that the shaded region is indeed the desired feasible region.

We now switch to the world where shoring is allowed and revisit the above two examples. The significant change is that the feasible curve that connects any two risky assets will be *infinitely* lengthened, and this will profoundly influence the shape of feasible region under this



(a) Example I



(b) Example II

Figure 2: Two typical shapes of feasible region for three risky assets when shorting is allowed (shaded in both grays), with comparison of feasible region under no shorting (shaded in dark gray only).

market condition. To be more precise, the new feasible regions, based on the data of previous two examples but with shorting permitted, are still *convex to the left* but become unbounded to the right and fully contain the (closed) feasible regions from no-shorting situation, as illustrated in Figure 2.

◊ “Convex to the left”的含义是，给定region里的任意两点，其连线不会超出region的左边边界. 造就这一特点的关键因素是我们在两资产讨论中得到的各种可能的feasible curve都是convex to the left(参考Topic1的Figures 4 or 5), 而此处的三资产亦或是更多资产所形成的feasible region正是由无数这样的两资产curve构造而成, 其结果一定是有一条curve成为左边边界(也叫minimum variance set,之前提到过).

Following the similar logic, we could easily imagine what the feasible region will look like when we add more risky assets. We summarize the important characteristics of the feasible region of *multiple* ($n \geq 3$) risky assets under both market conditions as follows:

- The feasible region, whether shorting is allowed or not, is convex to the left;
- the feasible region under no-shorting condition is closed with zigzag right boundary, while the feasible region with shorting available is unbounded to the right;
- the feasible region for short selling contains the one for no shorting, and the leftmost edges of these two regions may partially coincide (see Figure 2(b)) or may not (see Figure 2(a)).

◊ 可卖空条件下的feasible region一定包含不可卖空下的feasible region, 很好理解, 是因为前者的每个weight取值范围(整个$\mathbb{R}$)包含后者的取值范围(限定在[0,1]).

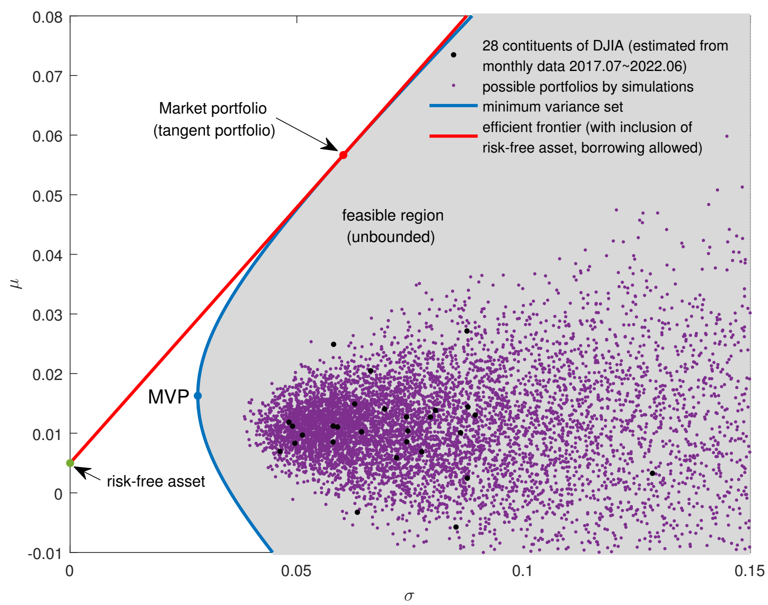


Figure 3: Feasible region and efficient frontier constructed by Matlab based on components of Dow Jones Industrial Average (shorting and borrowing are allowed).

5 Feasible region of a market: An example of Dow Jones Industrial Average

In this section, we will build up our own feasible region step by step, based on the real data from all the constituents of a stock-market index, namely, the Dow Jones Industrial Average (DJIA). We invoke Matlab to help us manipulate.[5] We assume both shorting and borrowing are allowed throughout the discussion. Figure 3 exhibits every detail mentioned in the steps below.

Step 0. We download monthly returns of all the components in DJIA ($n = 28$ hereafter[6], otherwise stated) from the vendor (like Bloomberg), ranging from July 2017 to June 2022 of recent five years ($m = 60$ hereafter, otherwise stated), and put them into a dataset $X \in \mathbb{R}^{m\times n}$. Notice that each column of X represents a time series of historical returns for a certain stock. The dataset X is used to estimate the mean vector $\mu \in \mathbb{R}^n$ and the covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ of random returns of all the stocks considered, namely,

$$\hat{\mu} = \frac{1}{m} X' \mathbb{1}_m, \ \hat{\Sigma} = \frac{1}{m-1}(X'X - m\hat{\mu}\hat{\mu}'), \tag{8}$$

as explained in Part I Preliminaries. Luckily, in Matlab the commands

```
mu_hat = mean(X,1)'; Sigma_hat = cov(X);
```

[5]The full code script and the dataset used will be published in Github/NickyFE/FinMath Topic2 branch.

[6]In general there should be 30 stocks inside DJIA but we delete two of them which have not been added into the index in the early days of our selected time period.

directly outputs the demanded results. Therefore, we could avoid coding (8) ourselves.

⚠ 这样偷懒直接用命令其实不好, 最好是掌握了估计背后的原理, 用起来才踏实.

Step 1. We plot each point of $(\hat{\sigma}_i, \hat{\mu}_i)$, $i = 1, 2, \ldots, n$ on the σ-μ graph, and then randomly sample some possible portfolios formed by these fundamental assets in order to identify a rough area that the feasible region would be. This is done by simulating some weights $w_k = (w_{k1}, w_{k2}, \ldots, w_{kn})' \in \mathbb{R}^n$, $k = 1, 2, \ldots, K$ such that each w_k satisfies the constraint $w_{k1} + w_{k2} + \cdots + w_{kn} = w_k' \mathbb{1}_n = 1$. Theoretically, the weight on every single asset could be $(-\infty, \infty)$ if shorting is available. As for practical sampling, we always set some boundaries such that it is first simulated by the Matlab command

```
w_k=-w_boundary+2*w_boundary*rand(n,1)
```

to get an n-dimensional vector `w_k` whose elements uniformly lie in [-`w_boundary`,`w_boundary`], and then *normalize* it by

```
w_k=w_k/(w_k'*ones(n,1))
```

that fulfils the constraint.[7] After we get the qualified weight `w_k`, we then compute the (estimated) mean and standard deviation of return of every possible portfolio through the commands

```
mu_k=w_k'*mu_hat,
sigma_k=sqrt(w_k'*Sigma_hat*w_k),
```

and plot them on the graph (see purple bullets in Figure 3).

Step 2. We now move to determine the feasible region of this market. Note that when the number of simulations, K, is as large as possible, we could see an approximated region that is occupied by those simulated portfolios. As we know, however, there are infinitely many portfolios in the feasible region, this approach is not attractive. On the other hand, if we follow the previous idea that we build up a region starting from a feasible curve of two individual assets, it is obviously time-consuming and worthless especially when we face many underlying assets. In fact, a more general way to tackle this problem is to find out the minimum variance set[8] first, and we know that the unbounded area contained by it is the feasible region we want. In Matlab, this could be (numerically) achieved by calling the optimization function

```
w_z = fmincon(obj_var,w_0,[],[],Aeq,beq),
```

[7]Note that if no-shorting condition is imposed, we could just simply do `w_k=rand(n,1)` and then pass to normalization.

[8]Recall that in Topic1 we have learnt that the minimum variance set collects portfolios with the lowest risk at each level of expected return. Graphically speaking, it is the left-most frontier of the feasible region formed by all given risky assets. In the simple case of two risky assets, it coincides with their feasible curve. Moreover, the left-most point on the minimum variance set is known as the minimum variance point (MVP).

after we define our objective function `obj_var=@(w) w'*Sigma_hat*w` which is to minimize the variance of portfolio return, subject to two constraints `w'ones(n,1)=1` and `w'mu_hat=z` where `z` is any target return we set, and these two constraints can be represented in a linear system of `w`, that is, `Aeq*w=beq`, where `Aeq=[ones(n,1)';mu_hat']` and `beq=[1;z]`. We also need to set an initial guess for `fmincon` to (numerically) solve this optimization problem, for instance an equal weight `w_0=1/n*ones(n,1)`. By varying the target return level `z`, we could obtain a series of optimal weights `w_z` for each `z` and finally trace out the minimum variance set formed by all (`sigma_z,mu_z`) (see blue curve in Figure 3). The entire unbounded area right to it is the feasible region of this market.[9] Besides, the MVP could be found out in a similar manner except that we only consider the constraint `w'ones(n,1)=1` in `fmincon`. Moreover, the upper branch of minimum variance set (i.e., the part when `mu_z`≥`mu_MVP`) is known as the efficient frontier for all risky assets.

Step 3. The last thing is to discover the efficient frontier when we add a risk-free asset F whose deterministic return is denoted by μ_F. As we discussed in Topic1, this can be done by maximizing the slope of the straight line that connects F and any possible portfolio formed by the original risky assets. We again leverage the function `fmincon` in Matlab based on the single constraint `w'ones(n,1)=1` but a new objective that calculates the *minus*[10] tangent (a.k.a. Sharpe ratio)

```
obj_tan=@(w) -(w'*mu_hat-mu_F)/sqrt(w'*Sigma_hat*w).
```

The optimal solution of this problem `w_T` constitutes the weight in tangent portfolio of this market, which is also known as the *market portfolio* in this case.

[9] From Figure 3 we can see that the simulated portfolios in this case are far from describing a correct feasible region.

[10] Because the function `fmincon` is trying to minimize the objective.

III. Markowitz Mean-variance Model

We revisit the entire studies in Topic1 and previous parts of Topic2 from the portfolio optimization point of view. Recall that we face a market with one risk-free asset with a deterministic return μ_F and n risky assets whose mean and covariance of random returns are given by μ and Σ (assume invertible), respectively. Our goal is threefold: i) Determine the risky portfolios that dominate others in the mean-variance sense (i.e., the minimum variance set); ii) find out the global MVP and hence the efficient frontier of all risky assets; iii) and finally discover the market portfolio and build up the efficient frontier that is achieved from all risky and riskless assets, a.k.a the capital market line (CML). Throughout the discussions below, we assume both shorting and borrowing are allowed.[11]

Note that these topics were initially established by Harry Markowitz in his series of seminal works from 1950s, who was awarded the Nobel Memorial Prize in Economic Sciences in 1990, and other famous scholars afterwards. See Further Readings at the end of this note for more details.

6 Minimum-variance set

The Markowitz mean-variance model seeks to find an optimal weight such that for a certain level of target expected return ("mean" side), the portfolio risk calculated by the variance of return ("variance" side) is minimized. Mathematically speaking, it would like to solve a portfolio optimization problem formulated by[12]

$$
(\mathcal{P}_z) \qquad \begin{array}{ll} \underset{w\in\mathbb{R}^n}{\text{minimize}} & \frac{1}{2}w'\Sigma w \\ \text{subject to} & w'\mathbb{1}_n = 1, \\ & w'\mu = z, \end{array}
$$

for some target expected return $z \in \mathbb{R}$.[13] It turns out to be a standard *quadratic* programming problem with equality constraints. We now introduce the Lagrange multipliers $\lambda, \gamma \in \mathbb{R}$ for each of the constraints and form the related Lagrangian

$$
L(w,\lambda,\gamma) = \frac{1}{2}w'\Sigma w - \lambda(w'\mathbb{1}_n - 1) - \gamma(w'\mu - z).
$$

After taking derivatives and setting the results to be zeros,

$$
\begin{aligned}
\frac{\partial L}{\partial w} &= \Sigma w - \lambda \mathbb{1}_n - \gamma\mu = 0, \\
\frac{\partial L}{\partial \lambda} &= -(w'\mathbb{1}_n - 1) = 0, \\
\frac{\partial L}{\partial \gamma} &= -(w'\mu - z) = 0,
\end{aligned}
$$

[11] No-shorting situation will be another mathematical story and we skip this in this note.

[12] The purpose of using $\frac{1}{2}$ in the objective is only because of the convenience when we take derivative on a *quadratic* function with symmetric Σ.

[13] Theoretically, the expected return of a portfolio or any asset could be negative, although we really do not want that.

we obtain the first-order optimality conditions for the unknowns (w, λ, γ),

$$\begin{cases} \Sigma w - \lambda \mathbb{1}_n - \gamma \mu = 0, \\ w' \mathbb{1}_n = 1, \\ w' \mu = z. \end{cases} \tag{9}$$

Obviously, the optimal solution differs when we change z. Suppose $(w^*(z), \lambda^*(z), \gamma^*(z))$ [for simplicity we write $(w_z^*, \lambda_z^*, \gamma_z^*)$] solve the above system of linear equations when the target expected return is set to be z, then we have

$$w_z^* = \lambda_z^* \Sigma^{-1} \mathbb{1}_n + \gamma_z^* \Sigma^{-1} \mu, \tag{10}$$

where λ_z^* and γ_z^* are scalars such that

$$\begin{cases} \lambda_z^* \mathbb{1}_n' \Sigma^{-1} \mathbb{1}_n + \gamma_z^* \mathbb{1}_n' \Sigma^{-1} \mu = 1, \\ \lambda_z^* \mu' \Sigma^{-1} \mathbb{1}_n + \gamma_z^* \mu' \Sigma^{-1} \mu = z. \end{cases} \tag{11}$$

The meaning of this solution is that the optimal weight w_z^* represents the minimum-variance portfolio we could achieve when our goal of portfolio's expected return is z, and this lowest variance is just computed through $(w_z^*)' \Sigma w_z^*$. By varying $z \in \mathbb{R}$, we can then trace out the minimum-variance set on the σ-μ graph, whose rigorous expression is actually given by

$$\{(\sqrt{(w_z^*)' \Sigma w_z^*},\ z) \mid \forall z \in \mathbb{R}\}.$$

Recall the leftmost boundary of the feasible region from the previous examples.

◊ 式(11)来自把(10)分别带入到(9)的后两个等式中. 在给定市场数据μ和Σ以及目标z时, 我们便可具体求得λ_z^*和γ_z^*以及w_z^*, 进而求得组合期望收益目标设为z时的组合收益率可达到的最小方差值, 即$(w_z^*)' \Sigma w_z^*$. 那此时组合期望收益率是多少? 当然是$(w_z^*)' \mu = z$.

7 Two-fund Theorem

The linearity of the equations in the optimality conditions (9) leads to a famous property of the minimum-variance set called *two-fund theorem*. More precisely, consider two solutions of (9), $(w_a, \lambda_a, \gamma_a)$ and $(w_b, \lambda_b, \gamma_b)$, corresponding to two different targets z_a and z_b, respectively. That is, assume for $j = a, b$,

$$\Sigma w_j - \lambda_j \mathbb{1}_n - \gamma_j \mu = 0, \tag{12}$$

$$w_j' \mathbb{1}_n = 1, \tag{13}$$

$$w_j' \mu = z_j.$$

Let us define a new triple $(w_\alpha, \lambda_\alpha, \gamma_\alpha)$ for some $\alpha \in \mathbb{R}$ such that

$$\begin{aligned} w_\alpha &= \alpha w_a + (1-\alpha) w_b, \\ \lambda_\alpha &= \alpha \lambda_a + (1-\alpha) \lambda_b, \\ \gamma_\alpha &= \alpha \gamma_a + (1-\alpha) \gamma_b, \end{aligned}$$

© 2022 Nicky老师

and it is easy to verify that this new triple also satisfies the optimality conditions (but for a new target return z_α as defined below). To see this, we just put it into the system and get

$$\begin{aligned}
&\Sigma w_\alpha - \lambda_\alpha \mathbb{1}_n - \gamma_\alpha \mu = \alpha(\Sigma w_a - \lambda_a \mathbb{1}_n - \gamma_a \mu) \\
&+ (1-\alpha)(\Sigma w_b - \lambda_b \mathbb{1}_n - \gamma_b \mu) = \alpha 0 + (1-\alpha)0 = 0, \\
&w_\alpha' \mathbb{1}_n = (\alpha w_a + (1-\alpha) w_b)' \mathbb{1}_n = \alpha 1 + (1-\alpha) 1 = 1, \\
&w_\alpha' \mu = (\alpha w_a + (1-\alpha) w_b)' \mu = \alpha z_a + (1-\alpha) z_b := z_\alpha.
\end{aligned}$$

The above results indicate that any minimum-variance portfolio can be duplicated (in terms of mean and variance) by the two funds w_a and w_b in a linear-mixture manner. In other words, investors who seek to invest on a certain minimum-variance portfolio are sufficient to invest on the combination of these two funds. On the other hand, w_a and w_b are actually free of choosing, as long as they themselves are the minimum-variance portfolios. Therefore, finding two particular funds is enough, and the easy way to achieve this is assigning $\lambda_a = 0$ and $\gamma_b = 0$. The corresponding weights under these two special cases are calculated in the following.

- By setting $\lambda_a = 0$ in (12), we first get

  $$w_a = \gamma_a \Sigma^{-1} \mu,$$

  and after the normalization [in order to fulfil (13)], we get

  $$w_a = \frac{\Sigma^{-1}\mu}{\mathbb{1}_n' \Sigma^{-1} \mu}; \tag{14}$$

- similarly for $\gamma_b = 0$ in (12), we finally obtain

  $$w_b = \frac{\Sigma^{-1}\mathbb{1}_n}{\mathbb{1}_n' \Sigma^{-1} \mathbb{1}_n}. \tag{15}$$

♢ Set $\lambda_a = 0$ and $\gamma_b = 0$并不是求得two funds的唯一选择, 但确是最容易求解的选择. 而此时z_a和z_b的值到底是多少并不重要, 因为我们只需求得两个特殊的weights即可.

Therefore, all the portfolios in the minimum-variance set can be represented by

$$\{w_\alpha \in \mathbb{R}^n \mid w_\alpha = \alpha w_a + (1-\alpha) w_b,\ \forall \alpha \in \mathbb{R}\}, \tag{16}$$

where w_a and w_b are given by (14) and (15), respectively. And the second way to describe the minimum-variance set on the graph is

$$\{(\sqrt{w_\alpha' \Sigma w_\alpha},\ w_\alpha' \mu) \mid \forall \alpha \in \mathbb{R}\}, \tag{17}$$

or based on

$$\begin{aligned}
&\mu_\alpha = w_\alpha' \mu = \alpha w_a' \mu + (1-\alpha) w_b' \mu,\ \alpha \in \mathbb{R}, \\
&\sigma_\alpha^2 = w_\alpha' \Sigma w_\alpha = (\alpha w_a + (1-\alpha) w_b)' \Sigma (\alpha w_a + (1-\alpha) w_b),\ \alpha \in \mathbb{R},
\end{aligned}$$

we have its explicit formula (check yourself that $(w_a - w_b)'\Sigma w_b = 0$)

$$\begin{aligned}\sigma_\alpha^2(\mu_\alpha) &= w_b'\Sigma w_b + 2\frac{\mu_\alpha - w_b'\mu}{w_a'\mu - w_b'\mu}(w_a - w_b)'\Sigma w_b + (\frac{\mu_\alpha - w_b'\mu}{w_a'\mu - w_b'\mu})^2(w_a - w_b)'\Sigma(w_a - w_b) \\ &= w_b'\Sigma w_b + \left(\frac{\mu_\alpha - w_b'\mu}{w_a'\mu - w_b'\mu}\right)^2 (w_a - w_b)'\Sigma(w_a - w_b),\ \mu_\alpha \in \mathbb{R}. \end{aligned} \tag{18}$$

And we can conclude two significant properties of the above concrete function: i) $\sigma_\alpha^2(\mu_\alpha) \geq w_b'\Sigma w_b$ for any $\mu_\alpha \in \mathbb{R}$ and the equality holds when $\mu_\alpha = w_b'\mu$ which is just the MVP; and ii) this parabolic curve is symmetric w.r.t. $\mu_\alpha = w_b'\mu$.

8 Minimum-variance point and efficient frontier of all risky assets

The minimum-variance point (MVP), as we have learnt before, is the *lowest*-risk portfolio (in terms of variance of portfolio return) that we can construct from all the given assets. From perspective of portfolio optimization, it can be solved from without considering the constraint of target expected return in $(\mathcal{P}_z)$, namely,

$$(\mathcal{MVP}) \qquad \begin{array}{ll} \underset{w\in\mathbb{R}^n}{\text{minimize}} & \frac{1}{2}w'\Sigma w \\ \text{subject to} & w'\mathbb{1}_n = 1. \end{array}$$

It is easy to obtain the optimal solution by considering the Lagrangian

$$L(w, \lambda) = \frac{1}{2}w'\Sigma w - \lambda(w'\mathbb{1}_n - 1),$$

and obtain the first-order optimality conditions

$$\Sigma w - \lambda\mathbb{1}_n = 0, \tag{19}$$

$$w'\mathbb{1}_n = 1. \tag{20}$$

From (19) we get $w = \lambda\Sigma^{-1}\mathbb{1}_n$ and put it into (20) we have $\lambda_{\text{MVP}} = 1/(\mathbb{1}_n'\Sigma^{-1}\mathbb{1}_n)$ and hence

$$w_{\text{MVP}} = \frac{\Sigma^{-1}\mathbb{1}_n}{\mathbb{1}_n'\Sigma^{-1}\mathbb{1}_n}.$$

Notice that w_{MVP} is same as w_b in (15), the second fund we build up in the previous two-fund theorem. It is not surprising to have this result, because when we set γ, the Lagrange multiplier for the constraint of target expected return, to be zero, it is just like we ignore this constraint and hence explore for the MVP directly.

Geometrically speaking, the upper portion of the minimum variance set (a parabolic curve on the σ-μ graph) is known as the efficient frontier of all risky assets considered. The turning point of the minimum variance set is just the MVP, whose expected return is given by

$$z_{\text{MVP}} = \mu' w_{\text{MVP}} = \frac{\mu'\Sigma^{-1}\mathbb{1}_n}{\mathbb{1}_n'\Sigma^{-1}\mathbb{1}_n}.$$

Therefore, the algebraic formula for the efficient frontier is given by

$$\{(\sqrt{(w_z^*)'\Sigma w_z^*},\ z) \mid \forall z \geq z_{\mathrm{MVP}}\}.$$

All the members of this set are known as the efficient portfolios with risky assets. Similar as the second expression for the minimum-variance set in (17), we could also represent the efficient frontier based on the two-fund theorem. However, this will depend on the relative positions of these two funds in terms of their expected returns. More specifically, if $\mu' w_a \geq \mu' w_b$ where w_b denotes the MVP, then the expression of the efficient frontier would be

$$\{(\sqrt{w_\alpha' \Sigma w_\alpha},\ w_\alpha' \mu) \mid \forall \alpha \geq 0,\ \text{if } \mu' w_a \geq \mu' w_b\}, \tag{21}$$

where w_α is given in (16) (note that the coefficient $(1-\alpha)$ is in front of the fund MVP); otherwise when $\mu' w_a < \mu' w_b$, it should be

$$\{(\sqrt{w_\alpha' \Sigma w_\alpha},\ w_\alpha' \mu) \mid \forall \alpha \leq 0,\ \text{if } \mu' w_a < \mu' w_b\}, \tag{22}$$

and in this case, the efficient frontier is formed by always short-selling the first fund w_a. But the easiest way to represent the efficient frontier is to use the explicit formula of the minimum-variance set in (18), as long as we set a proper domain for μ_α, that is,

$$\sigma_\alpha^2(\mu_\alpha) = w_b' \Sigma w_b + \left(\frac{\mu_\alpha - w_b'\mu}{w_a'\mu - w_b'\mu}\right)^2 (w_a - w_b)' \Sigma (w_a - w_b),\ \mu_\alpha \geq z_{\mathrm{MVP}}.$$

IV. Markowitz Mean-variance Model (Continued)

The last task of our discussions on mean-variance model is to investigate the situation that involves an additional risk-free asset. In the following, we provide three angles to study this problem.

9 Portfolio optimization with additional risk-free asset

Similar as the previous mean-variance problem $(\mathcal{P}_z)$, we now consider to add one more risk-free asset with deterministic return μ_F (usually positive value) and weight $w_F \in \mathbb{R}$ (borrowing is allowed). Therefore, the random return R_0 of the new portfolio P_0 from $n+1$ assets is

$$R_0 = w'R + w_F\mu_F,$$

where $R \in \mathbb{R}^n$ and $w \in \mathbb{R}^n$ are, respectively, the original returns and weights on all risky assets. The variance of this new return, however, would remain the same[14] as that of $w'R$. Therefore, we are facing a new portfolio optimization problem for a given expected return level $z \in \mathbb{R}$,

$$\begin{array}{llll} (\mathcal{Q}_z) & \underset{w\in\mathbb{R}^n, w_F\in\mathbb{R}}{\text{minimize}} & \frac{1}{2}w'\Sigma w \\ & \text{subject to} & w'\mathbb{1}_n + w_F = 1, \\ & & w'\mu + w_F\mu_F = z, \end{array}$$

If we solve w_F from the first constraint of $(\mathcal{Q}_z)$ and substitute the result into the second constraint, we would get a variant formulation

$$\begin{array}{llll} (\tilde{\mathcal{Q}}_z) & \underset{w\in\mathbb{R}^n, w_F=1-w'\mathbb{1}_n}{\text{minimize}} & \frac{1}{2}w'\Sigma w \\ & \text{subject to} & w'\bar{\mu} = z - \mu_F, \end{array}$$

where $\bar{\mu} = \mu - \mu_F\mathbb{1}_n$ is known as the *excess return* and we further assume that $\bar{\mu} \neq 0$. Once again, we apply the Lagrangian method to solve $(\tilde{\mathcal{Q}}'_z)$, namely, we introduce the Lagrange multiplier λ for the unique constraint left and form the Lagrangian

$$L(w,\lambda) = \frac{1}{2}w'\Sigma w - \lambda(w'\bar{\mu} - z + \mu_F).$$

The optimality equations are easily derived,

$$\begin{aligned} w &= \lambda\Sigma^{-1}\bar{\mu}, \\ w'\bar{\mu} &= z - \mu_F, \end{aligned}$$

[14] Recall that for a general random variable X and a constant a, $\operatorname{Var}(X+a) = \operatorname{Var}(X)$.

Therefore, the optimal multiplier is $\lambda_z^* = (z - \mu_F)/(\bar{\mu}'\Sigma^{-1}\bar{\mu})$, and the optimal weights on the risky and riskless portions would be

$$\begin{aligned} w_z^* &= (z-\mu_F)\frac{\Sigma^{-1}\bar{\mu}}{\bar{\mu}'\Sigma^{-1}\bar{\mu}}, \\ w_F^* &= 1 - (w_z^*)'\mathbb{1}_n = 1 - (z-\mu_F)\frac{\mathbb{1}_n'\Sigma^{-1}\bar{\mu}}{\bar{\mu}'\Sigma^{-1}\bar{\mu}}. \end{aligned} \tag{23}$$

◊ 注意, 正是因为假设了$\bar{\mu} \neq 0$, 才有了$\bar{\mu}'\Sigma^{-1}\bar{\mu} > 0$在分母的合法性, 而后者大于0是因为我们假设了$\Sigma$可逆也即正定. 后文我们也会谈到, $\mu \neq \mu_F \mathbb{1}_n$对tangent portfolio唯一性的重要作用.

And it is also easy to verify that the expected return of the current portfolio $P_0(z)$ with weights (w_z^*, w_F^*) is definitely equal to the settled target z,

$$\begin{aligned} \mu' w_z^* + \mu_F w_F^* &= (z-\mu_F)\frac{(\mu - \mu_F\mathbb{1}_n + \mu_F\mathbb{1}_n)'\Sigma^{-1}\bar{\mu}}{\bar{\mu}'\Sigma^{-1}\bar{\mu}} \\ &\quad + \mu_F - (z-\mu_F)\frac{\mu_F\mathbb{1}_n'\Sigma^{-1}\bar{\mu}}{\bar{\mu}'\Sigma^{-1}\bar{\mu}} \\ &= (z-\mu_F)\frac{\bar{\mu}'\Sigma^{-1}\bar{\mu}}{\bar{\mu}'\Sigma^{-1}\bar{\mu}} + (z-\mu_F)\frac{\mu_F\mathbb{1}_n'\Sigma^{-1}\bar{\mu}}{\bar{\mu}'\Sigma^{-1}\bar{\mu}} \\ &\quad + \mu_F - (z-\mu_F)\frac{\mu_F\mathbb{1}_n'\Sigma^{-1}\bar{\mu}}{\bar{\mu}'\Sigma^{-1}\bar{\mu}} \\ &= z. \end{aligned}$$

As usual, we could then build up the minimum-variance set among these $n+1$ assets,

$$\{(\sqrt{(w_z^*)'\Sigma w_z^*},\ z) \mid \forall z \in \mathbb{R}\}.$$

Moreover, if we compute explicitly the volatility of $P_0(z)$

$$\sigma_0(z) = \sqrt{(w_z^*)'\Sigma w_z^*} = \frac{|z-\mu_F|}{\sqrt{\bar{\mu}'\Sigma^{-1}\bar{\mu}}}, \tag{24}$$

we could find that the analytical formula for this set on the σ-μ plane is simply given by the inverse of (24), i.e.,

$$z(\sigma_0) = \mu_F \pm \sqrt{\bar{\mu}'\Sigma^{-1}\bar{\mu}}\sigma_0,\ \sigma_0 \geq 0,$$

which, as we could imagine from our previous experience, consists of two half-lines starting at the riskless point F and being *symmetric* w.r.t. the horizontal line across F.[15] Obviously, we have $\sigma_0(\mu_F) = 0$ when we set $z = \mu_F$ and at this time we invest merely on the risk-free asset which is also the MVP. Therefore, the efficient frontier with both risky and risk-free assets available (i.e., the upper branch of the minimum-variance set) could be represented by

$$\{(\sqrt{(w_z^*)'\Sigma w_z^*},\ z) \mid \forall z \geq \mu_F\}.$$

or explicitly given by

$$\begin{aligned} \sigma_0(z) &= \frac{z-\mu_F}{\sqrt{\bar{\mu}'\Sigma^{-1}\bar{\mu}}},\ z \geq \mu_F, \text{ or} \\ z(\sigma_0) &= \mu_F + \sqrt{\bar{\mu}'\Sigma^{-1}\bar{\mu}}\sigma_0,\ \sigma_0 \geq 0. \end{aligned}$$

[15] This statement can be conclude from two opposite slopes $\pm\sqrt{\bar{\mu}'\Sigma^{-1}\bar{\mu}}$.

10 One-fund theorem

If we treat the overall portfolio $P_0(z)$ as a combination of the risk-free asset F and a risky portfolio M, we know from the preceding section that w_F^* proportion of wealth is invested on F while the rest $(1-w_F^*)$ goes to M where n risky assets occupy w_z^* in $P_0(z)$. Then how about the composition inside M only? Apparently, the weights in this risky portfolio itself (when $w_F^* \neq 1$) can be easily computed through

$$w_M = \frac{w_z^*}{1-w_F^*} = \frac{\Sigma^{-1}\bar{\mu}}{\mathbb{1}_n'\Sigma^{-1}\bar{\mu}}, \tag{25}$$

which surprisingly is independent of z! In other words, if different investors want to invest on different minimum-variance portfolios from one riskless asset and n risky assets, all they need to care about is just determining the allocation between the risk-free asset F and a *fixed* risky portfolio M whose constituents within it is given by (25) (note that $w_M'\mathbb{1}_n = 1$). This crucial finding leads to an important principle in the mean-variance portfolio selection, namely, the *one-fund theorem* (a.k.a. mutual fund theorem, separation theorem). It says that any minimum-variance portfolio built up from both risky and risk-free assets can be directly constructed from the riskless asset and a single risky fund, which is called the *market portfolio*. Suppose we invest $w_F \in \mathbb{R}$ on F, then the minimum-variance portfolio could be represented by

$$\begin{pmatrix} w_F \\ (1-w_F)w_M \end{pmatrix},$$

where w_M given in (25) denotes the weights in market portfolio, which is only determined by the statistics of n risky assets and is independent of our preference. Accordingly, this portfolio's expected return and volatility are given by

$$\mu_0(w_F) = w_F\mu_F + (1-w_F)\mu_M = w_F\mu_F + (1-w_F)w_M'\mu, \tag{26}$$

$$\sigma_0^2(w_F) = (1-w_F)^2\sigma_M^2 = (1-w_F)^2 w_M'\Sigma w_M, \tag{27}$$

and hence the minimum-variance set can be written as

$$\{(\sigma_0(w_F),\ \mu_0(w_F)) \mid \forall w_F \in \mathbb{R}\}. \tag{28}$$

Besides, solving w_F from (26) and plugging it into (27) together with the expression of w_M in (25), we could obtain the explicit relationship between the mean and standard deviation,

$$\begin{aligned}
\sigma_0(\mu_0) &= \left|1 - \frac{\mu_0 - w_M'\mu}{\mu_F - w_M'\mu}\right| \sqrt{w_M'\Sigma w_M} \\
&= \left|\frac{\mu_0 - \mu_F}{\mu_F - \frac{(\mu - \mu_F\mathbb{1}_n + \mu_F\mathbb{1}_n)'\Sigma^{-1}\bar{\mu}}{\mathbb{1}_n'\Sigma^{-1}\bar{\mu}}}\right| \sqrt{\frac{\bar{\mu}'\Sigma^{-1}\Sigma\Sigma^{-1}\bar{\mu}}{(\mathbb{1}_n'\Sigma^{-1}\bar{\mu})^2}} \\
&= \frac{|\mu_0 - \mu_F||\mathbb{1}_n'\Sigma^{-1}\bar{\mu}|}{\bar{\mu}\Sigma^{-1}\bar{\mu}} \frac{\sqrt{\bar{\mu}\Sigma^{-1}\bar{\mu}}}{|\mathbb{1}_n'\Sigma^{-1}\bar{\mu}|} \\
&= \frac{|\mu_0 - \mu_F|}{\sqrt{\bar{\mu}\Sigma^{-1}\bar{\mu}}},
\end{aligned}$$

which is no-surprisingly same as (24). Therefore, we have learnt the second way to derive the minimum-variance set and hence the efficient frontier, which is from one-fund theorem perspective.

⚠ 如果想用类似(28)的方式表达efficient frontier, 要特别留意w_F的范围, 这完全取决于无风险资产F和marker portfolio M的相对位置: 理论上 M的期望收益率$w_M'\mu$可以大于也可以小于μ_F. 对比two-fund theorem中(21)和(22)的讨论. 我们会在(30)具体阐明这个细节.

11 Market portfolio via maximization on Sharpe ratio

As we have explained before, the feasible set between a risk-free asset and a risky asset, when both borrowing and shorting are available, consists of two half lines that starts at the same risk-free asset, and one of them passes the risky asset and the other one has an opposite slope thus geometrically speaking it is symmetric to the first half line w.r.t. the level of risk-free return (see Formula (29) and Figure 6 of Topic1 for exact details). This fact provides us a third way to find the minimum-variance portfolios which are built up from the riskless asset and all possible risky portfolios formed by n risky assets. Figure 4 exhibits the idea behind it. We know that the leftmost boundary of the feasible region of all risky assets, which is the minimum-variance set of all risky assets as well, is a parabolic curve[16] with the leftmost point known as the MVP. For a certain level of expected return z as illustrated in Figure 4, the portfolios formed by the risk-free asset F and a boundary portfolio P_z (purple lines) clearly dominates those formed by the same risk-free asset but a riskier portfolio $\tilde{P}_z$ that is located within the region (green lines). As for all possible boundary portfolios, it is easy to see that the tangent one would beat all the rest. Therefore, we conclude that any minimum-variance portfolio can be reached by investing on a combination of the risk-free asset and the tangent portfolio (this is exactly what the one-fund theorem has already told us but from a rigorous mathematical deduction). In order to determine this special risky portfolio, we could choose to maximize the slope of the half line that links the risk-free asset F and any risky portfolio P. It is also equivalent to maximize the Sharpe ratio of P. Therefore, we finally formulate the following problem,

$$(\mathcal{P}_M) \qquad \begin{aligned} &\underset{w\in\mathbb{R}^n}{\text{maximize}} \quad && \tan\theta = \frac{\mu_P-\mu_F}{\sigma_P} = \frac{w'\mu-\mu_F}{\sqrt{w'\Sigma w}} \\ &\text{subject to} && w'\mathbb{1}_n = 1. \end{aligned}$$

It is a fractional programming task at first glance, but luckily there exists a smart way to handle it. To see this, let us first denote the objective function by

$$\begin{aligned} f(w) &= (w'\mu-\mu_F)(w'\Sigma w)^{-\frac{1}{2}} \\ &= w'(\mu-\mu_F\mathbb{1}_n)(w'\Sigma w)^{-\frac{1}{2}}, \end{aligned}$$

where the second equality utilizes the constraint that $w'\mathbb{1}_n = 1$. Then we directly take first-order derivative on $f(w)$ to obtain the first-order optimality condition as if there is no

[16]See (18) to review its analytical expression. Notice that this curve is also symmetric w.r.t. the horizontal line at level $w'_{\text{MVP}}\mu$ (i.e., the expected return of MVP).

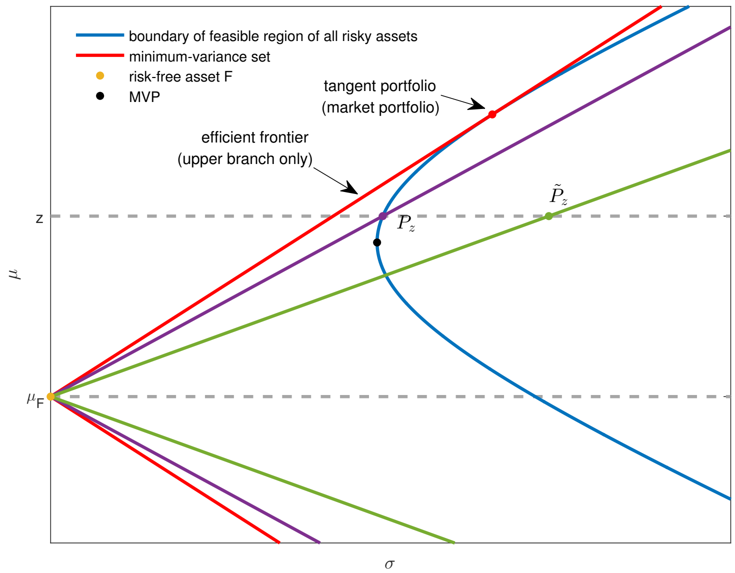


Figure 4: The idea of finding minimum-variance portfolios formed by a risk-free asset and a risky portfolio that is based on n risky assets (shorting and borrowing are allowed).

constraint at all. That is,

$$\begin{aligned}&\frac{\partial f}{\partial w} = (\mu - \mu_F \mathbb{1}_n)(w'\Sigma w)^{-\frac{1}{2}} - w'(\mu - \mu_F \mathbb{1}_n)(w'\Sigma w)^{-\frac{3}{2}}\Sigma w = 0,\\ &\Rightarrow\ (\mu - \mu_F \mathbb{1}_n) - w'(\mu - \mu_F \mathbb{1}_n)(w'\Sigma w)^{-1}\Sigma w = 0,\end{aligned}$$

and hence

$$w = \eta \Sigma^{-1}\bar{\mu}, \tag{29}$$

where again $\bar{\mu} = \mu - \mu_F \mathbb{1}_n$ (assume $\bar{\mu} \neq 0$) and η is a scalar that depends on w and is given by

$$\eta = \frac{w'\Sigma w}{w'\bar{\mu}}.$$

Note that the expression (29) is the sole structure of w that satisfies the first-order optimality condition. The remaining thing to do is to normalize (29) so that it fulfils the original constraint $w'\mathbb{1}_n = 1$. Therefore, the true optimal solution of $(\mathcal{P}_M)$ is given by

$$w_M = \frac{w}{\mathbb{1}_n' w} = \frac{\Sigma^{-1}\bar{\mu}}{\mathbb{1}_n'\Sigma^{-1}\bar{\mu}},$$

and this is exactly the market portfolio we have seen before. We could then express the minimum-variance set and also the efficient frontier as a linear combination between the riskless asset and this tangent portfolio (or market portfolio), same as what we did in the previous section on one-fund theorem.

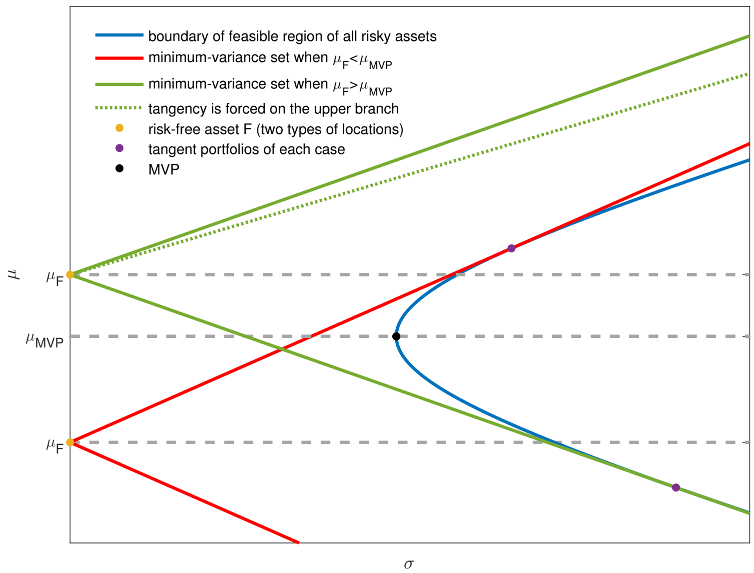


Figure 5: Two types of tangent portfolios according to the relationship between μ_F and μ_{MVP} (shorting and borrowing are allowed).

A remark on finding market portfolio

To end this topic, let us talk about a subtle issue which is related to the market portfolio but has been ignored by most of the existing textbooks.

Recall that when there is a risk-free asset apart from n risky assets, the weights of the market portfolio is given by $w_M = \Sigma^{-1}\bar{\mu}/(\mathbb{1}_n'\Sigma^{-1}\bar{\mu})$. This equation, and also other analytical formulas documented in the previous sections, are obtained based on the assumptions of

i) market conditions: both shorting and borrowing are allowed; and

ii) technical conditions: the covariance matrix Σ is invertible and the excess return $\bar{\mu} \neq 0$.

Figure 4 shows a geometrical fact that the market portfolio is just the tangent portfolio which appears on the *upper* branch of the risky boundary. This is what we could always see from the textbooks and note that in this case we have $\mu_F < \mu_{\text{MVP}}$.

However, we should bear in mind that it is *not* the unique location of market portfolio on the σ-μ plane when the above two assumptions are imposed. Theoretically, when $\mu_F > \mu_{\text{MVP}}$, w_M would be a tangent point on the *lower* branch of the boundary (see Figure 5 for a comparison). The classical textbooks do not discuss about this situation probably because they adopt by default $\mu_F < \mu_{\text{MVP}}$. In practice, however, it is possible to face a raised interest rate but a weaker equity market such that $\mu_F > \mu_{\text{MVP}}$ occurs (for example, during some period of year 2022). Hence, it is worthy to know this unusual phenomenon. The efficient frontier of this kind is actually constructed by short selling the market portfolio and invest more ($w_F \geq 1$) on the risk-free asset.[17] As for $\mu_F = \mu_{\text{MVP}}$, it is excluded by the assumption

[17] For instance, deposit more money into the bank to catch a higher interest rate than the equity return.

that $\bar{\mu} \neq 0$.[18] However, we could easily imagine that if it was the case, then there would be two tangent points simultaneously since the risky boundary is just symmetric w.r.t. the horizontal level μ_{MVP} (see (18) for the symmetry).

Therefore, the representation of efficient frontier based on the minimum-variance set (28) in Section 10 is distinguished into the following two cases:

$$\begin{aligned}&\{(\sigma_0(w_F),\ \mu_0(w_F)) \mid \forall w_F \leq 1 \text{ if } \mu_F < \mu_{\text{MVP}}\}; \text{ or}\\ &\{(\sigma_0(w_F),\ \mu_0(w_F)) \mid \forall w_F \geq 1 \text{ if } \mu_F > \mu_{\text{MVP}}\}.\end{aligned} \tag{30}$$

Last but not least, readers may get confused that in the model of maximization on the Sharpe ratio, we do maximize the slope but why do we seemingly get some negative slope when the tangent portfolio happens at the lower branch of the curve? Remember that the minimum-variance set consists of two half lines with opposite slope values, and due to the allowance of short selling and borrowing, the optimization problem $(\mathcal{P}_M)$ actually aims to find a risky portfolio such that the upper slope (or equivalently, the absolute value of lower slope) is maximized. Figure 5 also exhibits a half line that is forced to be tangent on the upper branch of the risky boundary when $\mu_F > \mu_{\text{MVP}}$ (the green dotted line, the tangency happens out of the scope of current figure) and we can easily see that it is dominated by the real efficient frontier (the upper half of the green solid line).[19]

Further readings

Chapter 6 and the reference papers in the end of this chapter, *Investment Science*, Luenberger, D. G., Oxford University Press, 2013.

[18] This is because when $\mu = \mu_F \mathbb{1}_n$, we have $\mu_{\text{MVP}} = \mu' w_{\text{MVP}} = \mu_F \mathbb{1}_n' \Sigma^{-1} \mathbb{1}_n / (\mathbb{1}_n' \Sigma^{-1} \mathbb{1}_n) = \mu_F$. However, we already assume that $\mu \neq \mu_F \mathbb{1}_n$.

[19] Apparently, when shorting is forbidden, there will be a unique market portfolio tangent on the upper branch of the risky boundary. But notice that in this situation the feasible region (and hence the boundary) will be changed as well, compared to the case when shorting is allowed.