Modern Portfolio Theory

Revisited through Modern Calculus

Nikolai Nowaczyk *

This script revisits the classical Modern Portfolio Theory (MPT) and Capital Asset Pricing Model (CAPM) by Markowitz, Sharpe et al. using modern calculus. The treatment aims to connect the core financial concepts and economic intuition with a modern mathematical treatment. We introduce the Gaussian model of asset returns and show how this leads to diversification effects. Exploiting those we then give various fully self-contained derivations of the efficient frontier using constrained optimization and Lagrange multipliers. We then show how the introduction of a risk-free asset changes the picture to the Capital Allocation Line (CAL) and relate this to the key concept of a Sharpe ratio. Finally, we use these insights to derive the Capital Asset Pricing Model (CAPM) and shed some light on the passive vs. active investment theory.

*Contact: mail@nikno.de

Contents

1 **Introduction** . . . 3

2 **Asset & Returns** . . . 4

3 **The Gaussian Model** . . . 5
- 3.1 Portfolio Returns & Volatilities . . . 6
- 3.2 Diversification . . . 8

4 **Modern Portfolio Theory (MPT)** . . . 9
- 4.1 Risk Minimization . . . 9
- 4.2 The Efficient Frontier . . . 11
- 4.3 Return Maximization . . . 14
- 4.4 Markowitz Duality . . . 16
- 4.5 Two Mutual Fund Theorem . . . 18

5 **Capital Allocation Line (CAL)** . . . 19
- 5.1 One Fund Theorem (Risk Minimization) . . . 20
- 5.2 One Fund Theorem (Return Maximization) . . . 22
- 5.3 Tangency Portfolio . . . 23
- 5.4 Sharpe Ratio . . . 25
- 5.5 Two Fund Separation Theorem . . . 26
- 5.6 Sharpe Ratio Maximization . . . 26

6 **Capital Asset Pricing Model (CAPM)** . . . 28
- 6.1 Assets and Portfolios . . . 28
- 6.2 CAPM Formulations . . . 30
- 6.3 The Beta & (Non)systematic Risk . . . 31
- 6.4 CAPM and Simple Regression . . . 32
- 6.5 Abnormal Returns . . . 33
- 6.6 The Alpha & Performance Measurement . . . 34

7 **Enduring Insights & Practical Limitations** . . . 35

8 **Appendix** . . . 35

List of Symbols . . . 37

1. Introduction

Modern Portfolio Theory (MPT) is based on the landmark paper [4] published by Henry Markowitz in 1952, for which he was awarded the Nobel Memorial Prize in Economic Sciences in 1990. It addresses on of the biggest and most heatedly debated question of portfolio management: Which stocks should one select? The key contribution is the insight that one should consider risks as well as returns. By taking volatilities and correlations into account, one can make use of powerful diversification effects. Maximizing the portfolio returns given an acceptable level of risk, or, equivalently, minimizing portfolio risks given a target return, leads to the famous concept of an efficient frontier, see Fig. 1.

In his paper [4], Markowitz introduces these key financial concepts and mathematically proves the case for $d = 3$ assets geometrically. In this script we first introduce some basic notions around asset prices and returns Section 2, introduce the fundamental Gaussian model of asset returns in Section 3 and illustrate diversification can be mathematically explained in that model. In the core Section 4, we give fully self-contained proofs of Markowitz' results for the general case of d assets using modern calculus, in particular the theory of Lagrange multipliers to solve constrained optimization problems. We give a summary of the results needed in Section 8. These results apply to markets of risky assets only, which is not realistic. Thus, in Section 5 we repeat the study for markets that have a risk-free asset in addition leading to the famous Capital Allocation Line (CAL) and the notion of a Sharpe ratio. Finally, in Section 6 we use the results to derive how individual asset returns relate to market returns, discuss the implications for active vs. passive investing and how CAPM can be used to assess performance of asset managers.

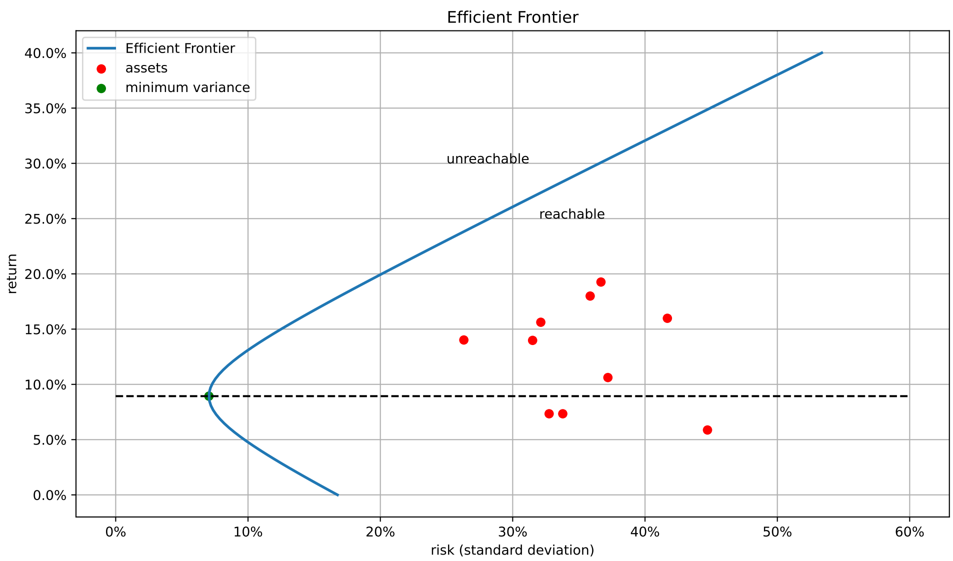


Figure 1: Efficient Frontier

2. Asset & Returns

In this section we summarize some basic notions around total returns, rate of returns and their relationship to asset prices and cashflows.

Definition 2.1 (returns)**.** Let $(S_t)_{0 \le t \le n}$ be a time series of asset prices of length $n+1$. The time series $(R'_t)_{1 \le t \le n}$ of length n defined by

$$R'_t := S_t - S_{t-1} \tag{2.1}$$

is called the total *capital gains/losses* of the asset S. Let $(C_t)_{1 \le t \le n}$ be the time series of cashflows of the asset S (e.g. dividends, coupons, etc.). Then

$$R_t := R'_t + C_t = S_t - S_{t-1} + C_t. \tag{2.2}$$

are called the *total holding period returns.* $\diamondsuit$

In practice cashflows are an important component of the returns. The theory is a bit smoother though for assets that don't generate cashflows.

Lemma 2.2 (prices from total returns)**.** Let R_t be the total returns of an asset with initial price S_0 and cashflows $(C_t)_{1 \le t \le n}$. Then for any $1 \le t \le n$,

$$S_t = S_0 + \sum_{i=1}^{t} R'_i = S_0 + \sum_{i=1}^{t} R_i - \sum_{i=1}^{t} C_i \tag{2.3}$$

In particular, any two assets price time series with the same capital gains/losses only differ by a constant (given by the difference in their initial values). $\diamondsuit$

Proof. The first equation can be seen via induction over t. Clearly, for $t = 0$, this equation holds. To see the implication $t \mapsto t+1$, we calculate

$$S_{t+1} \overset{(2.1)}{=} R'_{t+1} + S_t \overset{(2.3)}{=} R_{t+1} + S_0 + \sum_{i=1}^{t} R'_t + R'_{t+1} = S_0 + \sum_{i=1}^{t+1} R'_t.$$

To see the second equation, we notice that

$$\sum_{i=1}^{t} R'_i \overset{(2.2)}{=} \sum_{i=1}^{t} (R_i - C_i) = \sum_{i=1}^{t} R_i - \sum_{i=1}^{t} C_i. \qquad \square$$

The total returns of assets can be on very different orders of magnitudes and denominated in very different currencies and units, which makes them hard to compare. Thus, we introduce the rate of return.

Definition 2.3 (rate of return)**.** Let $(S_t)_{0 \le t \le n}$ be a time series of asset prices with total returns $(R_t)_{1 \le t \le n}$. Then is the *holding period rate of return* $(r_t)_{1 \le t \le n}$ is defined by:

$$r_t := \frac{R_t}{S_{t-1}} = \frac{S_t - S_{t-1} + C_t}{S_{t-1}} = \underbrace{\frac{S_t - S_{t-1}}{S_{t-1}}}_{=:r'_t} + \underbrace{\frac{C_t}{S_{t-1}}}_{=:c_t} = r'_t + c_t \tag{2.4}$$

The quantity r'_t is the *rate of capital gains/losses* and c_t is often called *dividend yield / coupon rate / etc.* $\diamondsuit$

Notice that in the absence of cashflows, $r_t = r'_t$ is simply the percentage change of the asset price. The terminology is not entirely coherent in the literature. Often, r_t are simply called "the returns".

Lemma 2.4 (prices from rate of returns)**.** Let $(r_t)_{1\le t\le n}$ be the rate of return of an asset with initial price S_0 and dividend yields $(c_t)_{1\le t\le n}$. Then for any $0 \le t \le n$,

$$S_t = S_0 \prod_{i=1}^{t} (1 + r'_i) = S_0 \prod_{i=1}^{t} (1 + r_i - c_i) \tag{2.5}$$

In particular, any two assets with the same rate of return have a constant ratio. $\Diamond$

Proof. The first equation can be seen by induction over t. Clearly, this holds for $t = 0$. To see the implication $t \to t+1$, we first use the definitions to obtain

$$r'_{t+1} = \frac{R'_{t+1}}{S_t} = \frac{S_{t+1} - S_t}{S_t} = \frac{S_{t+1}}{S_t} - 1.$$

Thus

$$S_{t+1} = (1 + r'_{t+1})S_t = (1 + r'_{t+1})S_0 \prod_{i=1}^{t} (1 + r'_i) = S_0 \prod_{i=1}^{t+1} (1 + r'_i),$$

which implies the first equation. The second equation follows from (2.4). $\square$

Remark 2.5. This formula can be used to express the returns $r'_{0,t}$ over n holding periods:

$$r'_{0,t} := \frac{S_n - S_0}{S_0} = \frac{S_0 \prod_{i=1}^{t} (1 + r_i) - S_0}{S_0} = \prod_{i=1}^{t} (1 + r'_i) - 1$$

In particular, if the asset goes down by $p\%$ in the first period and then down by the same $p\%$ in the second:

$$r'_{0,2} = (1 - p)(1 + p) - 1 = 1 - p^2 - 1 = -p^2 \neq 0. \qquad \Diamond$$

3. The Gaussian Model

Modern portfolio theory is based on the Gaussian model of asset returns. Throughout we make the following assumptions:

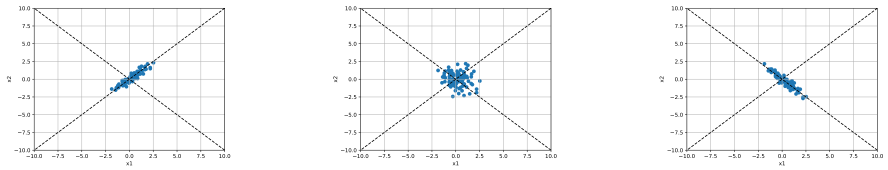
Assumption 3.1 (Modern Portfolio Theory Model Assumptions)**.**

- There exists a fixed universe of d assets $S = (S^{(1)}, \ldots, S^{(d)})$, which can be bought and sold in any fractional amount, positive and negative at zero transaction costs.
- At any fixed point in time, the return vector $r = (r^{(1)}, \ldots, r^{(d)})$ describing the rate of return of these assets over a fixed horizon (e.g. 1Y) is a multivariate Gaussian[1]random variable, i.e.

  $$r \sim \mathcal{N}(\mu, \Sigma)$$

  with a mean vector $\mu \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$.

[1] see e.g. `https://en.wikipedia.org/wiki/Multivariate_normal_distribution`



(a) positive correlation (b) zero correlation (c) negative correlation

Figure 2: Impact of Correlation

- We make the technical assumptions that
  - The returns $\mu \in \mathbb{R}^d$ and $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^d$ are linearly independent.
  - The covariance matrix Σ is *positive-definite*[2]. $\diamondsuit$

Remark 3.2. Instead of working with a covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$ given by

$$\Sigma_{ij} = \mathrm{Cov}[r^{(i)}, r^{(j)}] = \mathbb{E}[(r^{(i)} - \mu_i)(r^{(j)} - \mu_j)] \in \mathbb{R}$$

one can also work with a correlation[3] matrix $C \in \mathbb{R}^{d \times d}$ given by

$$\rho_{ij} := \frac{\Sigma_{ij}}{\sigma_i \sigma_j} = \frac{\mathrm{Cov}[r^{(i)}, r^{(j)}]}{\sqrt{\mathrm{Var}[r^{(i)}]}\sqrt{\mathrm{Var}[r^{(j)}]}} \in [-1, 1]$$

and a vector $\sigma^2 = (\sigma_1^2, \ldots, \sigma_d^2)$ of variances, or, alternatively of volatilities $\sigma = (\sigma_1, \ldots, \sigma_d)$. Denoting by $D(\sigma)$ the diagonal matrix with the volatilities, one can also write this as

$$\Sigma = D(\sigma) C D(\sigma). \qquad \diamondsuit$$

We do not aim to introduce multivariate Gaussians in this script from scratch, but want to convey one very important intuition. Say, we are considering two assets, then positive correlation means their returns tend to move together, see Fig. 2a. If they are uncorrelated the returns are independently from each other, see Fig. 2b. If they are negatively correlated, the returns tend to move in opposite directions, see Fig. 2c.

3.1. Portfolio Returns & Volatilities The problem of forming a portfolio P can hence be reduced of specifying a relative weight each asset has in the portfolio.

Definition 3.3 (weight vector)**.** A *weight vector* is a vector $w \in \mathbb{R}^d$ such that

$$\sum_{i=1}^{d} w_i = 1. \qquad \diamondsuit$$

Remark 3.4. In many other mathematical contexts, one requires a weight to be non-negative, i.e. $w_i \geq 0$, which then implies $w_i \leq 1$. We explicitly allow weights to be negative. The financial interpretation of this is that we allow for short-selling. $\diamondsuit$

Weights are relative weights, i.e. in percent. In order to execute a plan to actually form a portfolio by buying/selling the assets, the relative weights have to be translated into absolute units.

[2] see e.g. https://en.wikipedia.org/wiki/Definite_matrix

[3] see e.g. https://en.wikipedia.org/wiki/Correlation

Definition 3.5 (units vector)**.** Let $S = (S^{(1)}, \dots, S^{(d)})$ be a market of d assets of value $S_0^{(i)}$ at $t = 0$ and $w = (w_1, \dots, w_d)$ be a weight vector. For any investment amount A, we define a corresponding vector $u = (u_1, \dots, u_d)$ of *units* related to the portfolio value P_0 at $t = 0$ via

$$u_i := \frac{w_i A}{S_0^{(i)}}. \qquad \lozenge$$

By construction, the value P_0 of the portfolio at $t = 0$ satisfies

$$P_0 = \sum_{i=1}^{d} u_i S_0^{(i)}.$$

Lemma 3.6. Assume no cashflows. Let $P = \sum_{i=1}^{d} u_i S_0^{(i)}$ be a portfolio value at $t = 0$. At $t = 1$ (day/year/etc.) the total returns of the portfolio satisfy

$$R_P := P_1 - P_0 = \sum_{i=1}^{d} u_i R^{(i)} \tag{3.1}$$

and the rates of return of the portfolio satisfy

$$r_P := \frac{R_P}{P_0} = \sum_{i=1}^{d} w_i r^{(i)}. \tag{3.2}$$

$\lozenge$

Proof. We calculate for the total returns

$$R_P = \sum_{i=1}^{d} u_i S_1^{(i)} - \sum_{i=1}^{d} u_i S_0^{(i)} = \sum_{i=1}^{d} u_i (S_1^{(i)} - S_0^{(i)}) = \sum_{i=1}^{d} u_i R^{(i)}$$

and the rate of return

$$\begin{aligned} r_P = \frac{R_P}{P_0} &= \frac{1}{P_0} \sum_{i=1}^{d} u_i (S_1^{(i)} - S_0^{(i)}) = \frac{1}{P_0} \sum_{i=1}^{d} u_i ((1 + r^{(i)}) S_0^{(i)} - S_0^{(i)}) \\ &= \frac{1}{P_0} \sum_{i=1}^{d} \frac{w_i A}{S_0^{(i)}} S_0^{(i)} (1 + r^{(i)} - 1) = \sum_{i=1}^{d} w_i r^{(i)} \end{aligned} \qquad \square$$

Remark 3.7. The (3.2) is of key importance to the model. It implies that the portfolio returns are a weighted sum of the asset returns. The financial implications are that they can be worked out entirely in percent without any reference to any currency. The mathematical implications are that portfolio returns are linear combination of the Gaussian asset returns, which enables an efficient theoretical treatment. $\lozenge$

Theorem 3.8. Assume that $r = (r^{(1)}, \dots, r^{(d)}) \sim \mathcal{N}(\mu, \Sigma)$ is a Gaussian vector of asset returns. Then the portfolio returns $r_P = \sum_{i=1}^{d} w_i r^{(i)} = w^\top r$ are also Gaussian satisfying

$$r_P \sim \mathcal{N}(\mu_P, \sigma_P^2),$$

where the portfolio returns are given by

$$\mu_P := \mathbb{E}[r_P] = w^\top \mu = \sum_{i=1}^{d} w_i \mu_i \tag{3.3}$$

and the portfolio variance is given by

$$\sigma_P^2 := \mathrm{Var}[r_P] = w^\top \Sigma w = \sum_{i,j} w_i w_j \Sigma_{ij} = \sum_{i,j=1}^{d} w_i w_j \sigma_i \sigma_j \rho_{ij} \tag{3.4}$$

$$= \sum_i w_i^2 \sigma_i^2 + 2 \sum_{i<j} w_i w_j \sigma_i \sigma_j \rho_{ij}. \tag{3.5}$$

◊

This claim follows from the fact that linear combination of Gaussian random variables are Gaussian.

3.2. Diversification The following is the mathematical foundation of a key principle of portfolio optimization.

Theorem 3.9 (diversification)**.** In the Gaussian model with d assets it is always true that

$$\mu_P = \sum_{i=1}^{d} w_i \mu_i, \qquad \sigma_P \leq \sum_{i=1}^{d} w_i \sigma_i.$$

◊

Proof. The first equation follows from the linearity of expectations. For the second, we calculate

$$\sigma_P^2 = \sum_i w_i^2 \sigma_i^2 + 2 \sum_{i<j} w_i w_j \sigma_i \sigma_j \underbrace{\rho_{ij}}_{\leq 1} \leq \sum_i w_i^2 \sigma_i^2 + 2 \sum_{i<j} w_i w_j \sigma_i \sigma_j$$

$$= \sum_{i,j} w_i w_j \sigma_i \sigma_j = \Big(\sum_{i=1} w_i \sigma_i \Big)^2.$$

□

This theorem has profound implications. It states that the volatility of the portfolio is always less (or equal) than the weighted volatilities of the assets. The counter-intuitive consequence of this is that you can reduce your risk by taking risks. The proof resolves this paradox: If the new asset has low, maybe even negative, correlation with the rest of the portfolio assets, the total risk is reduced. This is the mathematical reasoning behind diversification, the widely spread principle of not putting all your eggs in one basket. It also shows that portfolio volatility is driven not only by the asset volatilities, but even more so by their correlation structure.

Remark 3.10. In the simplified case where all assets weights are equal, i.e. $w_i := \frac{1}{d}$, we have

$$\mu_P = \frac{1}{d} \sum_{i=1}^{d} \mu_i = \bar{\mu}$$

$$\sigma_P^2 \approx \frac{1}{d} \bar{\sigma}^2 + \frac{d-1}{d} \bar{\sigma}^2 \bar{\rho} \xrightarrow[d \to \infty]{} \bar{\sigma}^2 \bar{\rho},$$

where $\bar{\mu}$, $\bar{\sigma}$, $\bar{\rho}$ are the average return, risk and correlation. This shows that if correlations are small enough, these can yield a big reduction in portfolio risk. This equation is quite

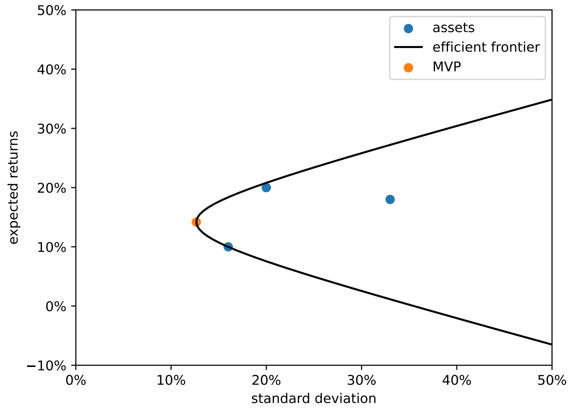


Figure 3: Efficient Frontier

famous, but the first $\approx$ is only a rule of thumb. The other steps can be rigorously proven:

$$\begin{aligned}\sigma_P^2 &= \sum_i w_i^2\sigma_i^2 + 2\sum_{i<j} w_iw_j\sigma_i\sigma_j\rho_{ij} = \frac{1}{d^2}\sum_i \sigma_i^2 + \frac{2}{d^2}\sum_{i<j}\sigma_i\sigma_j\rho_{ij} \\ &\approx \frac{1}{d^2}\sum_i \bar\sigma^2 + \frac{2}{d^2}\sum_{i<j}\bar\sigma^2\bar\rho = \frac{1}{d}\bar\sigma^2 + \frac{2}{d^2}\frac{d(d-1)}{2}\bar\sigma^2\bar\rho \\ &= \frac{1}{d}\bar\sigma^2 + \frac{d-1}{d}\bar\sigma^2\bar\rho\end{aligned}$$

$\diamondsuit$

Remark 3.11 (two asset example)**.** For $d = 2$ we can conveniently express the first asset weight as $w := w_1$ and the second as $w_2 = 1-w$ and thus obtain a neat formula for portfolio return and variance in this case:

$$\begin{aligned}\mu_P &= w\mu_1 + (1-w)\mu_2 \\ \sigma_P^2 &= w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2w(w-1)\sigma_1\sigma_2\rho\end{aligned}$$

$\diamondsuit$

4. Modern Portfolio Theory (MPT)

4.1. Risk Minimization We now prove the first version of the Markowitz optimization: Given the universe of d assets, the distribution $r \sim \mathcal{N}(\mu, \Sigma)$ of their returns, and a target return μ_P we want to achieve with a portfolio, what are the optimal portfolio weights $w^* \in \mathbb{R}^d$ in order to minimize the volatility σ_P^* of the portfolio but still achieving the expected target returns μ_P? The following theorem gives the answer.

Theorem 4.1 (Markowitz Variance Minimization)**.** For any chosen target returns μ_P, the problem

$$\text{minimize } w \mapsto \frac{1}{2}w^\top\Sigma w \quad \text{subject to } w^\top\mathbf{1} = \sum_{i=1}^d w_i = 1 \quad \text{and} \quad \mu^\top w = \sum_{i=1}^d w_i\mu_i = \mu_P,$$

has a unique solution

$$w_\triangledown = w_\triangledown(\mu_P) = \frac{C - B\mu_P}{D}\Sigma^{-1}\mathbf{1} + \frac{A\mu_P - B}{D}\Sigma^{-1}\mu, \tag{4.1}$$

where

$$A := \mathbf{1}^\top \Sigma^{-1}\mathbf{1} \qquad B := \mathbf{1}^\top \Sigma^{-1}\mu \qquad C := \mu^\top\Sigma^{-1}\mu, \qquad D := AC - B^2 \tag{4.2}$$

are all scalar. The variance pertaining to $w_\triangledown$ is given by

$$(\sigma_P^\triangledown)^2 = \sigma_P^\triangledown(\mu_P)^2 = \frac{A\mu_P^2 - 2B\mu_P + C}{D}. \tag{4.3}$$

$\diamondsuit$

Proof. Define the objective function $f : \mathbb{R}^d \to \mathbb{R},\ w \mapsto \frac{1}{2}w^\top\Sigma w$. Its derivative is given by

$$\nabla f(w) = w^\top\Sigma \in \mathbb{R}^{1\times d}.$$

Define the two constrains

$$g_1 : \mathbb{R}^d \to \mathbb{R},\ w \mapsto w^\top\mathbf{1} \qquad\qquad g_2 : \mathbb{R}^d \to \mathbb{R},\ w \mapsto w^\top\mu,$$

which have derivatives

$$\nabla g_1(w) = \mathbf{1}^\top \in \mathbb{R}^{1\times d} \qquad\qquad \nabla g_2(w) = \mu^\top \in \mathbb{R}^{1\times d}.$$

By Theorem 8.1 there exists $\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2$ such that a critical point $w_\triangledown$ satisfies

$$\nabla f(w_\triangledown) = (\lambda_1, \lambda_2)\begin{pmatrix}\nabla g_1(w_\triangledown)\\ \nabla g_2(w_\triangledown)\end{pmatrix} = \lambda_1\nabla g_1(w_\triangledown) + \lambda_2\nabla g_2(w_\triangledown).$$

Plugging in the definitions, we obtain the Euler Lagrange equation

$$(w_\triangledown)^\top\Sigma = \lambda_1\mathbf{1}^\top + \lambda_2\mu^\top \implies \Sigma w_\triangledown = \lambda_1\mathbf{1} + \lambda_2\mu. \implies w_\triangledown = \Sigma^{-1}(\lambda_1\mathbf{1} + \lambda_2\mu). \tag{4.4}$$

Plugging this into the constraints gives two more equations

$$\begin{aligned} 1 = \mathbf{1}^\top w_\triangledown &= \lambda_1\underbrace{\mathbf{1}^\top\Sigma^{-1}\mathbf{1}}_{=A} + \lambda_2,\underbrace{\mathbf{1}^\top\Sigma^{-1}\mu}_{=B} = \lambda_1 A + \lambda_2 B \\ \mu_P = \mu^\top w_\triangledown &= \lambda_1\underbrace{\mu^\top\Sigma^{-1}\mathbf{1}}_{=B} + \lambda_2\underbrace{\mu^\top\Sigma^{-1}\mu}_{=C} = \lambda_1 B + \lambda_2 C. \end{aligned}$$

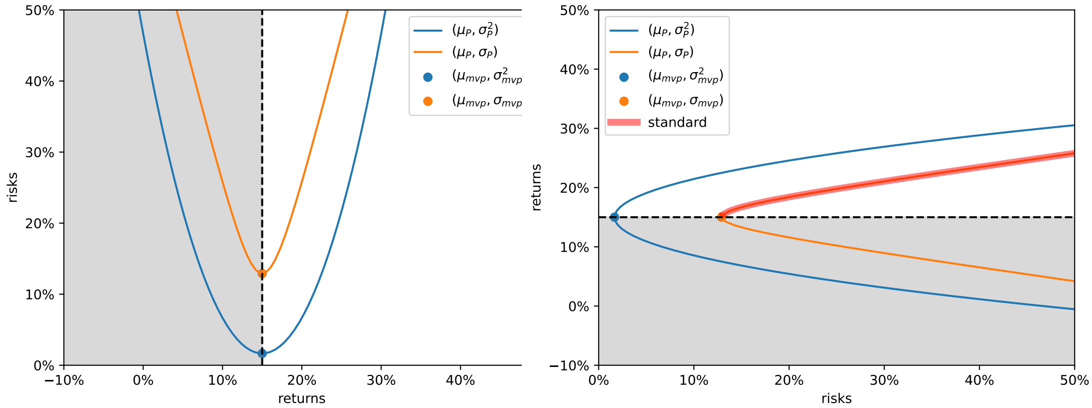
The key insight is that $A, B, C \in \mathbb{R}$, thus these satisfy a 2 x 2 system of linear equations

$$\begin{pmatrix}A & B\\ B & C\end{pmatrix}\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix} = \begin{pmatrix}1\\ \mu_P\end{pmatrix}.$$

The determinant of that system is $D := AC - B^2$. Because Σ^{-1} is positive-definite the Cauchy-Schwarz inequality[4] for the scalar product $\langle\_,\_\rangle_{\Sigma^{-1}}$ induced by Σ^{-1}states

$$B^2 = |\mathbf{1}^\top\Sigma^{-1}\mu|^2 = |\langle\mathbf{1}, \mu\rangle_{\Sigma^{-1}}|^2 \le \langle\mathbf{1},\mathbf{1}\rangle_{\Sigma^{-1}}\langle\mu,\mu\rangle_{\Sigma^{-1}} = AC \implies D = AC - B^2 \ge 0$$

[4]see e.g. https://en.wikipedia.org/wiki/Cauchy%E2%80%93Schwarz_inequality



(a) Return-Risk (b) Risk-Return

Figure 4: Parametrizations of Efficient Frontier

with equality if and only if μ and $\mathbf{1}$ are linearly dependent. Since we have assumed independence in Assumption 3.1, we conclude $D > 0$. Thus, we have the explicit formula

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} A & B \\ B & C \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \mu_P \end{pmatrix} = \frac{1}{D} \begin{pmatrix} C & -B \\ -B & A \end{pmatrix} \begin{pmatrix} 1 \\ \mu_P \end{pmatrix} = \frac{1}{D} \begin{pmatrix} C - B\mu_P \\ -B + A\mu_P \end{pmatrix}$$

and the optimal weights are explicitly given by

$$\begin{aligned} w_\triangledown &\overset{(4.4)}{=} \Sigma^{-1}(\lambda_1 \mathbf{1} + \lambda_2 \mu) = \frac{1}{D}\Sigma^{-1}((C - B\mu_P)\mathbf{1} + (A\mu_P - B)\mu) \\ &= \frac{C - B\mu_P}{D}\Sigma^{-1}\mathbf{1} + \frac{A\mu_P - B}{D}\Sigma^{-1}\mu. \end{aligned}$$

For the variance, we calculate

$$\begin{aligned} (\sigma_P^\triangledown)^2 &= w_\triangledown^\top \Sigma w_\triangledown = w_\triangledown^\top \Big(\frac{C - B\mu_P}{D}\mathbf{1} + \frac{A\mu_P - B}{D}\mu \Big) \\ &= \Big(\frac{C - B\mu_P}{D}\mathbf{1}^\top \Sigma^{-1} + \frac{A\mu_P - B}{D}\mu^\top \Sigma^{-1} \Big) \Big(\frac{C - B\mu_P}{D}\mathbf{1} + \frac{A\mu_P - B}{D}\mu \Big) \\ &= \frac{1}{D^2}\Big((C - B\mu_P)^2 A + 2(C - B\mu_P)(A\mu_P - B)B + (A\mu_P - B)^2 C \Big) \\ &= \frac{1}{D^2}\Big(AC^2 - 2ABC\mu_P + AB^2\mu_P^2 + 2ABC\mu_P - 2B^2C - 2AB^2\mu_P^2 + 2B^3\mu_P \\ &\qquad + A^2C\mu_P^2 - 2ABC\mu_P + B^2C \Big) \\ &= \frac{1}{D^2}\Big(\mu_P^2(AB^2 - 2AB^2 + A^2C) + \mu_P(-2ABC + 2B^3 - 2ABC + 2ABC) \\ &\qquad + AC^2 - 2B^2C + B^2C \Big) \\ &= \frac{1}{D^2}\Big(\mu_P^2(A^2C - AB^2) - 2\mu_P(ABC - B^3) + AC^2 - B^2C \Big) \\ &= \frac{1}{D^2}\Big(AD\mu_P^2 - 2\mu_P BD + CD \Big) = \frac{A\mu_P^2 - 2B\mu_P + C}{D}. \end{aligned}$$

□

4.2. The Efficient Frontier One way of thinking about Theorem 4.1 is that we have found a way to assign to any target return μ_P a portfolio $w_\triangledown$ that achieves these returns and is of minimal variance $(\sigma_P^{\nabla})^2$. We can thus think of this assignment as a function. This function is very famous and has a special name.

Definition 4.2 (efficient frontier)**.** The function

$$\mu_P \mapsto \sigma_P^{\nabla}(\mu_P)^2 := \frac{A\mu_P^2 - 2B\mu_P + C}{D}, \tag{4.5}$$

where A, B, C, D are given in (4.2), is called *Efficient Frontier.* $\Diamond$

Notice that the coefficients A, B, C, D depend on the market model (μ, Σ). Thus, this function essentially solves the investment problem. For any given target returns, it tells us the risks in form of the variance we have to accept. In Fig. 4a we have plotted the graph of this function. One should remark that this function is in fact so famous that there exist slightly different forms.

Variance vs. Volatility When plotting the function (4.5), we plot returns μ_P against variances σ_P^2, which is called *return-variance parametrization.* The result is a parabola, which is a very nice easily tractable function mathematically. In most mathematical theory, working with variances is generally easier than working with volatilities. However, for practical purposes, recall that the returns are given in percent, e.g. $\mu_P = 20\%$. Analogously, the volatilities are also given in percent, e.g. $\sigma_P = 10\%$, which means that returns and volatilities can be compared very nicely. The variance would be $\sigma_P^2 = 0.1^2 = 0.01$, which is a much smaller number, but of course doesn't correspond to lower risk. Because of these practical considerations, many people prefer to plot the risks as volatilities, i.e. standard deviations, rather than variances, i.e. μ_P against σ_P, which is also shown in Fig. 4a. As we can see, the shape of the graph using volatilities is qualitatively similar, basically just a bit more narrow, than the graph using variances.

Global Minimal Variance The efficient frontier in its return-variance parametrization is a parabola. Like all parabolas, this means it has a global minimum, which we can calculate, see Corollary 4.3. This is a remarkable result of profound practical relevance: In presence of risky assets, we can reduce risk through diversification and portfolio construction, but not below an absolute threshold. This threshold σ_{mvp}^2 depends only on the market covariance Σ and also leads to the lowest returns μ_{mvp} associated to this level of risk. Now, if we aim for higher returns, i.e. target $\mu_P > \mu_{\mathrm{mvp}}$, then we have to live with higher levels of risk. That makes sense. However, targeting lower returns $\mu_P < \mu_{\mathrm{mvp}}$, also increases the risk from a purely mathematical point of view as evident from the plot of the parabola. This is not financially meaningful though. Once we know the minimal level of risk σ_{mvp} and the associated returns μ_{mvp}, any rational investor would either accept that or target higher returns accepting higher risks. Nobody would reasonably target lower returns than those associated with the lowest risk at the cost of higher risks. For that reason, the left half of the parabola in Fig. 4a is often deleted.

Corollary 4.3 (Global Minimum Variance Portfolio (MVP))**.** Assume that the returns of a portfolio follow the d-dimensional normal distribution $\mathcal{N}(\mu, \Sigma)$. The problem

$$\text{minimize } \frac{1}{2} w^\top \Sigma w \qquad\qquad \text{subject to } \mathbf{1}^\top w = \sum_{i=1}^{d} w_i = 1.$$

has a unique solution satisfying

$$w_{\mathrm{mvp}} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^\top \Sigma^{-1}\mathbf{1}} = \frac{1}{A}\Sigma^{-1}\mathbf{1}, \quad \sigma^2_{\mathrm{mvp}} = \frac{1}{\mathbf{1}^\top \Sigma^{-1}\mathbf{1}} = \frac{1}{A}, \quad \mu_{\mathrm{mvp}} = \frac{\mathbf{1}^\top \Sigma^{-1}\mu}{\mathbf{1}^\top \Sigma^{-1}\mathbf{1}} = \frac{B}{A} \tag{4.6}$$

$\diamondsuit$

There are two ways to prove this: We can either solve this via Lagrangians with one condition dropped or we can use the result of the previous theorem and minimize over all target returns. We give both proofs here.

Proof. Using Lagrangians is a much easier version of the previous technique: The Lagrangian is now given by

$$\Sigma w_{\mathrm{mvp}} = \lambda \mathbf{1} \Longrightarrow w_{\mathrm{mvp}} = \lambda \Sigma^{-1}\mathbf{1}\,.$$

Thus, plugging this into the only constraint gives

$$1 = \mathbf{1}^\top w_{\mathrm{mvp}} = \lambda \mathbf{1}^\top \Sigma^{-1}\mathbf{1} \Longrightarrow \lambda = \frac{1}{\mathbf{1}^\top \Sigma^{-1}\mathbf{1}},$$

thus

$$\begin{aligned} w_{\mathrm{mvp}} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^\top \Sigma^{-1}\mathbf{1}} &\Longrightarrow \sigma^2_{\mathrm{mvp}} = w_{\mathrm{mvp}}^\top \Sigma w_{\mathrm{mvp}} = \lambda^2 \mathbf{1}^\top \Sigma^{-1}\mathbf{1} = \frac{1}{\mathbf{1}^\top \Sigma^{-1}\mathbf{1}} = \frac{1}{A}, \\ \Longrightarrow \mu_{\mathrm{mvp}} = \mu^\top w_{\mathrm{mvp}} = \frac{\mu^\top \Sigma^{-1}\mathbf{1}}{\mathbf{1}^\top \Sigma^{-1}\mathbf{1}} &= \frac{B}{A}. \end{aligned}$$

$\square$

Proof. This result can also be obtained by using (4.3)

$$\sigma_P^{\triangledown}(\mu_P)^2 = \frac{A(\mu_P)^2 - 2B\mu_P + C}{D}.$$

and finding its minimal value:

$$0 \overset{!}{=} \frac{\partial \sigma_P^{\triangledown}(\mu_P)^2}{\partial \mu_P}(\mu_{\mathrm{mvp}}) = \frac{2A\mu_{\mathrm{mvp}} - 2B}{D} \Longrightarrow A\mu_{\mathrm{mvp}} = B \Longrightarrow \mu_{\mathrm{mvp}} = \frac{B}{A}.$$

This results in

$$\sigma^2_{\mathrm{mvp}} = \frac{A\frac{B^2}{A^2} - 2B\frac{B}{A} + C}{D} = \frac{\frac{B^2}{A} - 2\frac{B^2}{A} + C}{D} = \frac{\frac{AC - B^2}{A}}{D} = \frac{1}{A}$$

$\square$

Risk-Return Parametrization It turns out that flipping the axes on the graph of Fig. 4a, which leads to Fig. 4b also gives an interesting result. This now plots the return we have to target once we accept a given level of risk. This is actually much closer to practical portfolio construction. The mathematical problem that there are now two points on the parabola above a risk σ_P conveniently disappears if we again delete the half of the parabola that is not financially meaningful anyway. This parametrization is called the *risk-return parametrization* and by assigning μ_P to volatilities σ_P we arrive at a graphical representation that is most useful for practical purposes, which is why this is probably the most common. We will use this as the *standard parametrization* of the risk-return space from here on.

Of course the process of flipping the axes of the plot and deleting the lower part of the resulting function can be described mathematically as well.

Corollary 4.4. Let $(\mu_P, \sigma_P^\triangledown)$ be a point on the (financially meaningful half) of the efficient frontier (4.5), i.e.

$$(\sigma_P^\triangledown)^2 = \frac{A\mu_P^2 - 2B\mu_P + C}{D}.$$

Then conversely,

$$\mu_P = \frac{B}{A} + \frac{\sqrt{D}}{A}\sqrt{A(\sigma_P^\triangledown)^2 - 1}. \tag{4.7}$$

◊

Proof. We simply solve the quadratic equation for μ_P:

$$\begin{aligned}(\sigma_P^\triangledown)^2 = \frac{A\mu_P^2 - 2B\mu_P + C}{D} &\iff \frac{D(\sigma_P^\triangledown)^2}{A} = \mu_P^2 - 2\frac{B}{A}\mu_P + \frac{C}{A} \\ &\iff \mu_P^2 - 2\frac{B}{A}\mu_P + \frac{B^2}{A^2} = \frac{D(\sigma_P^\triangledown)^2 - C}{A} + \frac{B^2}{A^2} \\ &\iff \left(\mu_P - \frac{B}{A}\right)^2 = \frac{AD(\sigma_P^\triangledown)^2 - AC + B^2}{A^2} = D\frac{A(\sigma_P^\triangledown)^2 - 1}{A^2}\end{aligned}$$

We know that $D \geq 0$ and $A > 0$. From Corollary 4.3 we also know that

$$(\sigma_P^\triangledown)^2 \geq (\sigma_{\text{mvp}}^\triangledown)^2 = \frac{1}{A} \implies A(\sigma_P^\triangledown)^2 - 1 \geq 0.$$

Hence, mathematically this quadratic equation has the two solutions

$$\mu_P = \frac{B}{A} \pm \frac{\sqrt{D}}{A}\sqrt{A(\sigma_P^\triangledown)^2 - 1},$$

and the bigger one is financially meaningful. (In case A $(\sigma_P^\triangledown)^2 - 1 = 0$ we are at the minimum and we only have one solution anyway.) □

4.3. Return Maximization In (4.7) we have expressed the target returns μ_P as a function of the minimized variance $(\sigma_P^\triangledown)^2$. However, we could also have approached this topic the other way around: Given a target level of risk σ_P, maximize the returns. This approaching the problem of portfolio optimization is equally if not even more legitimate from a practical financial point of view. Therefore, we also present this alternative derivation in the following.

Theorem 4.5 (Markowitz Return Maximization)**.** Assuming the Gaussian model $r \sim \mathcal{N}(\mu, \Sigma)$. Then for any target variance σ_P^2, the problem

$$\text{maximize } w \mapsto w^\top \mu \qquad \text{subject to } w^\top \mathbf{1} = \sum_{i=1}^d w_i = 1 \qquad \text{and} \qquad w^\top \Sigma w = \sigma_P^2,$$

has a unique solution

$$w_\triangle = w_\triangle(\sigma_P) = \frac{\sqrt{A\sigma_P^2 - 1}}{\sqrt{D}}\Sigma^{-1}\mu + \frac{\sqrt{D} - B\sqrt{A\sigma_P^2 - 1}}{A\sqrt{D}}\Sigma^{-1}\mathbf{1}. \tag{4.8}$$

The returns pertaining to $w_\triangle$ are given by

$$\mu_P^\triangle = \mu_P^\triangle(\sigma_P) = \frac{\sqrt{D}}{A}\sqrt{A\sigma_P^2 - 1} + \frac{B}{A}. \tag{4.9}$$

◊

Proof. We rewrite the second constraint to the equivalent condition $\frac{1}{2}w^\top \Sigma w = \frac{1}{2}\sigma_P^2$. Now, we again apply Theorem 8.1 and obtain

$$\mu = \eta_1 \mathbf{1} + \eta_2 \Sigma w_\triangle,$$

for some Lagrange multiplier $\eta = (\eta_1, \eta_2)$. If $\eta_2 = 0$, then μ and $\mathbf{1}$ would be linearly dependent contradicting Assumption 3.1, thus $\eta_2 \neq 0$ and this is equivalent to

$$w_\triangle = \frac{1}{\eta_2}\Sigma^{-1}(\mu - \eta_1 \mathbf{1}).$$

Plugging this into the first constraint, we obtain

$$1 = \mathbf{1}^\top w_\triangle = \frac{1}{\eta_2}(B - \eta_1 A).$$

We can re-arrange this to

$$\eta_1 = \frac{B - \eta_2}{A}.$$

Plugging $w_\triangle$ into the second constraint gives

$$\begin{aligned}
\eta_2^2\sigma_P^2 &= \eta_2^2(w_\triangle)^\top \Sigma w_\triangle = (\mu - \eta_1 \mathbf{1})^\top \Sigma(\mu - \eta_1 \mathbf{1}) \\
&= C - 2\eta_1 B + \eta_1^2 A = C - 2B\frac{B-\eta_2}{A} + \Big(\frac{B-\eta_2}{A}\Big)^2 A \\
&= \frac{AC + -2B^2 + 2B\eta_2 + B^2 - 2B\eta_2 + \eta_2^2}{A} \\
&= \frac{AC\eta_2^2 - B^2}{A} = \frac{AC - B^2 + \eta_2^2}{A} = \frac{D + \eta_2^2}{A},
\end{aligned}$$

thus

$$A\eta_2^2\sigma_P^2 - \eta_2^2 = D \Longrightarrow \eta_2^2 = \frac{D}{A\sigma_P^2 - 1}.$$

Plugging these expressions of the Lagrange multipliers back in, we obtain

$$\begin{aligned}
w_\triangle &= \frac{1}{\eta_2}\Sigma^{-1}(\mu - \eta_1 \mathbf{1}) = \frac{1}{\eta_2}\Sigma^{-1}\Big(\mu - \frac{B - \eta_2}{A}\mathbf{1}\Big) \\
&= \frac{1}{\eta_2}\Sigma^{-1}\Big(\mu - \frac{B}{A}\mathbf{1}\Big) + \frac{1}{A}\Sigma^{-1}\mathbf{1} \\
&= \frac{\sqrt{A\sigma_P^2 - 1}}{\sqrt{D}}\Sigma^{-1}\Big(\mu - \frac{B}{A}\mathbf{1}\Big) + \frac{1}{A}\Sigma^{-1}\mathbf{1} \\
&= \frac{\sqrt{A\sigma_P^2 - 1}}{\sqrt{D}}\Sigma^{-1}\mu + \frac{\sqrt{D} - B\sqrt{A\sigma_P^2 - 1}}{A\sqrt{D}}\Sigma^{-1}\mathbf{1}
\end{aligned}$$

To obtain the maximized returns, we calculate

$$\begin{aligned}
\mu_P^\triangle = \mu^\top w_\triangle &= \frac{\sqrt{A\sigma_P^2 - 1}}{\sqrt{D}}\mu^\top\Sigma^{-1}\Big(\mu - \frac{B}{A}\mathbf{1}\Big) + \frac{1}{A}\mu^\top\Sigma^{-1}\mathbf{1} \\
&= \frac{\sqrt{A\sigma_P^2 - 1}}{\sqrt{D}}\Big(\mu^\top\Sigma^{-1}\mu - \frac{B}{A}\mu^\top\Sigma^{-1}\mathbf{1}\Big) + \frac{1}{A}\mu^\top\Sigma^{-1}\mathbf{1} \\
&= \frac{\sqrt{A\sigma_P^2 - 1}}{\sqrt{D}}\Big(C - \frac{B^2}{A}\Big) + \frac{B}{A} = \frac{\sqrt{A\sigma_P^2 - 1}}{\sqrt{D}}\Big(\frac{AC - B^2}{A}\Big) + \frac{B}{A} \\
&= \frac{\sqrt{A\sigma_P^2 - 1}}{\sqrt{D}A}D + \frac{B}{A} = \frac{\sqrt{D}}{A}\sqrt{A\sigma_P^2 - 1} + \frac{B}{A}.
\end{aligned}$$

□

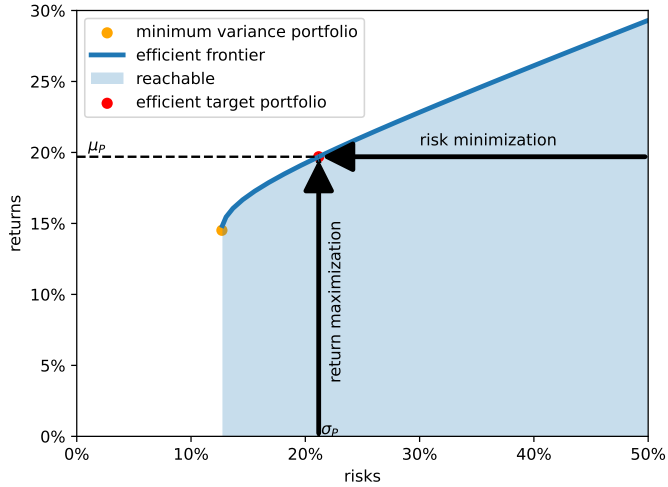


Figure 5: Risk Minimization vs Return Maximization

Just like in Corollary 4.4, we can also write the efficient frontier the other way around.

Corollary 4.6. The target variance σ_P^2 can conversely be written as a function of the maximal returns $\mu_P^\triangle$ via

$$\sigma_P^2 = \frac{A(\mu_P^\triangle)^2 - 2B\mu_P^\triangle + C}{D}. \tag{4.10}$$

$\diamondsuit$

Proof. We calculate

$$\mu_P^\triangle = \frac{\sqrt{D}}{A}\sqrt{A\sigma_P^2 - 1} + \frac{B}{A} \implies \frac{A\mu_P^\triangle - B}{\sqrt{D}} = \sqrt{A\sigma_P^2 - 1} \implies \frac{(A\mu_P^\triangle - B)^2}{D} = A\sigma_P^2 - 1,$$

thus

$$\begin{aligned}\sigma_P^2 &= \frac{1}{A}\Big(\frac{(A\mu_P^\triangle - B)^2}{D} + 1\Big) = \frac{1}{A}\Big(\frac{A^2(\mu_P^\triangle)^2 - 2AB\mu_P^\triangle + B^2 + AC - B^2}{D}\Big)\\ &= \frac{A(\mu_P^\triangle)^2 - 2B\mu_P^\triangle + C}{D}.\end{aligned}$$

$\square$

4.4. Markowitz Duality We have now derived two version of a Markowitz optimal portfolio: One by setting target returns and minimizing the variance, one by setting target variance and maximizing the returns. This raises the natural question if these two approaches are equivalent. It turns out the answer to this question is: yes, even though this might not be immediately obvious from the formulas Eqs. (4.1) and (4.8) for $w_\triangledown$ and $w_\triangle$, but maybe more so if we take a look at efficient frontier again, see Fig. 5. That is because the former is written as a function of the target variance where the later is written as a function of a target return. As we will show below, these weights do agree when appropriately re-written.

And that is not a coincidence. There is a deep mathematical reason for this revealed in convex analysis called *dualization*. Under broad conditions one can replace a *primal problem*, e.g. minimize variance given returns, by a *dual problem*, e.g. maximize returns given variance, and it turns out both have the same solution, see e.g. [2, Chapter XII]. However, it is often the case that one problem is much easier to solve in practice than the other. In this case though, we already solved the primal problem directly in Theorem 4.1 and it turns out the dual problem is equally easily solved. We prove this below in Theorem 4.5 directly, i.e. without using abstract principles from convex analysis.

Theorem 4.7 (Markowitz Duality)**.** Let μ_P be the target returns for the portfolio. Let $w_\triangledown = w_\triangledown(\mu_P)$ be the weights of the portfolio with minimal variance $(\sigma_P^\triangledown)^2 = \sigma_P^\triangledown(\mu_P)^2$ pertaining to μ_P from Theorem 4.1. Set the target variance $\sigma_P^2 := \sigma_P^\triangledown(\mu_P)^2$ and let $w_\triangle = w_\triangle(\sigma_P^\triangledown)$ be the weights of portfolio with maximal returns $\mu_P^\triangle = \mu_P^\triangle(\sigma_P^\triangledown)$ pertaining to $\sigma_P^2 = (\sigma_P^\triangledown)^2$ from Theorem 4.5. Then

$$w_\triangle(\sigma_P^\triangledown) = w_\triangledown(\mu_P^\triangle), \qquad\qquad \mu_P = \mu_P^\triangle(\sigma_P^\triangledown).$$

Conversely, let σ_P be a target variance and let $w_\triangle = w_\triangle(\sigma_P)$ be the weights of the portfolio with maximal returns $\mu_P^\triangle = \mu_P^\triangle(\sigma_P)$ pertaining to σ_P from Theorem 4.5. Set the target returns to $\mu_P := \mu_P^\triangle$ and let $w_\triangledown = w_\triangledown(\mu_P^\triangle)$ be the weights of the portfolio with minimal variance $(\sigma_P^\triangledown)^2 = \sigma_P^\triangledown(\mu_P^\triangle)$ pertaining to $\mu_P := \mu_P^\triangle$ Theorem 4.1. Then

$$w_\triangledown(\mu_P^\triangle) = w_\triangle(\sigma_P) \qquad\qquad \sigma_P = \sigma_P^\triangledown(\mu_P^\triangle). \qquad \diamondsuit$$

Proof. Note that from the first statement we have already shown that $\mu_P = \mu_P^\triangle$ by (4.7). From the second claim we have already shown $\sigma_P = \sigma_P^\triangledown$ by (4.10). It remains to show the claim on the portfolio weights. This can be shown in two ways: It follows indirectly from Theorems 4.1 and 4.5 because the weights of the optimal portfolio are unique.
Alternatively, one can also show this via direct computation: Notice that by (4.7)

$$\mu_P^\triangle = \frac{\sqrt{D}}{A}\sqrt{A(\sigma_P^\triangledown)^2 - 1} + \frac{B}{A} \iff \sqrt{A(\sigma_P^\triangledown)^2 - 1} = \frac{A\mu_P^\triangle - B}{\sqrt{D}} \tag{4.11}$$

Plugging this into (4.8), we obtain

$$\begin{aligned}
w_\triangle(\sigma_P^\triangledown) &\overset{(4.8)}{=} \frac{\sqrt{A(\sigma_P^\triangledown)^2 - 1}}{\sqrt{D}}\Sigma^{-1}\mu + \frac{B\sqrt{A(\sigma_P^\triangledown)^2 - 1} - \sqrt{D}}{A\sqrt{D}}\Sigma^{-1}\mathbf{1} \\
&= \frac{\frac{A\mu_P^\triangle - B}{\sqrt{D}}}{A\sqrt{D}}\Sigma^{-1}\mu + \frac{B\frac{A\mu_P^\triangle - B}{\sqrt{D}} - \sqrt{D}}{A\sqrt{D}}\Sigma^{-1}\mathbf{1} \\
&= \frac{\frac{A\mu_P^\triangle - B}{\sqrt{D}}}{A\sqrt{D}}\Sigma^{-1}\mu + \Big(\frac{AB\mu_P^\triangle - B^2}{AD} - \frac{1}{A}\Big)\Sigma^{-1}\mathbf{1} \\
&= \frac{\frac{A\mu_P^\triangle - B}{\sqrt{D}}}{A\sqrt{D}}\Sigma^{-1}\mu + \frac{AB\mu_P^\triangle - B^2 - D}{AD}\Sigma^{-1}\mathbf{1} \\
&= \frac{\frac{A\mu_P^\triangle - B}{\sqrt{D}}}{A\sqrt{D}}\Sigma^{-1}\mu + \frac{AB\mu_P^\triangle - AC}{AD}\Sigma^{-1}\mathbf{1} \overset{(4.1)}{=} w_\triangledown(\mu_P^\triangle).
\end{aligned}$$

Conversely, notice that since

$$\begin{aligned}
\frac{\sqrt{D} - B\sqrt{A(\sigma_P^\triangledown)^2 - 1}}{A\sqrt{D}} &\overset{(4.11)}{=} \frac{\sqrt{D} - B\frac{A\mu_P^\triangle - B}{\sqrt{D}}}{A\sqrt{D}} = \frac{1}{A} - \frac{B}{A}\frac{A\mu_P^\triangle - B}{D} \\
&= \frac{D - AB\mu_P^\triangle + B^2}{AD} = \frac{AC - AB\mu_P^\triangle}{AD} = \frac{C - B\mu_P^\triangle}{D},
\end{aligned} \tag{4.12}$$

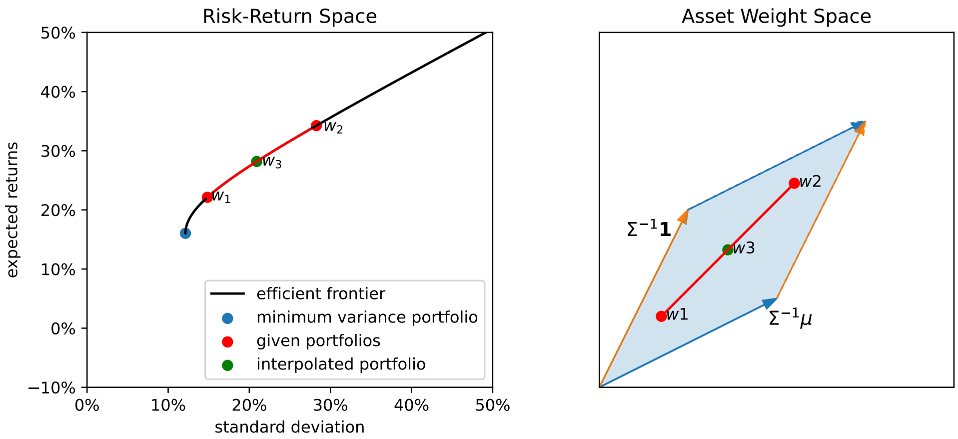


Figure 6: Two Mutual Fund Theorem

we obtain

$$
\begin{aligned}
w_\nabla(\mu_P^\triangle) &\overset{(4.1)}{=} \frac{C - B\mu_P^\triangle}{D}\Sigma^{-1}\mathbf{1} + \frac{A\mu_P^\triangle - B}{D}\Sigma^{-1}\mu \\
&\overset{(4.11)}{=} \frac{\sqrt{A(\sigma_P^\nabla)^2 - 1}}{\sqrt{D}}\Sigma^{-1}\mu + \frac{C - B\mu_P^\triangle}{D}\Sigma^{-1}\mathbf{1} \\
&\overset{(4.12)}{=} \frac{\sqrt{A\sigma_P^2 - 1}}{\sqrt{D}}\Sigma^{-1}\mu + \frac{\sqrt{D} - B\sqrt{A\sigma_P^2 - 1}}{A\sqrt{D}}\Sigma^{-1}\mathbf{1} \overset{(4.8)}{=} w_\triangle(\sigma_P^\nabla).
\end{aligned}
$$
□

4.5. Two Mutual Fund Theorem Recall from (4.1) that the portfolio weights w_∇ of the variance minimal portfolio are given by

$$
w_\nabla = \frac{C - B\mu_P}{D}\Sigma^{-1}\mathbf{1} + \frac{A\mu_P - B}{D}\Sigma^{-1}\mu,
$$

i.e. they are a linear combination of just two vectors $\Sigma^{-1}\mathbf{1}$ and $\Sigma^{-1}\mu$ both multiplied by scalars, which are affine linear functions of the target returns μ_P. This mathematical fact can be exploited to show the practically interesting financial fact that once we have found two distinct efficient portfolios on the frontier and know their optimal weights, we can obtain any other point on the efficient frontier, by just linearly combining the weights, see Fig. 6.

Theorem 4.8 (Two Mutual Fund Theorem)**.** Let $w_\nabla^{(1)}$ and $w_\nabla^{(2)}$ be the optimal weights pertaining to two variance-minimal portfolios on the efficient frontier with target returns $\mu_P^{(1)}$ and $\mu_P^{(2)}$. Let $\mu_P^{(3)}$ be any other target return. Then the optimal weights $w_\nabla^{(3)}$ of the third portfolio are a linear combination of the first two:

$$
w_\nabla^{(3)} = \frac{\mu_P^{(3)} - \mu_P^{(2)}}{\mu_P^{(1)} - \mu_P^{(2)}} w_\nabla^{(1)} + \frac{\mu_P^{(1)} - \mu_P^{(3)}}{\mu_P^{(1)} - \mu_P^{(2)}} w_\nabla^{(2)}. \tag{4.13}
$$
◊

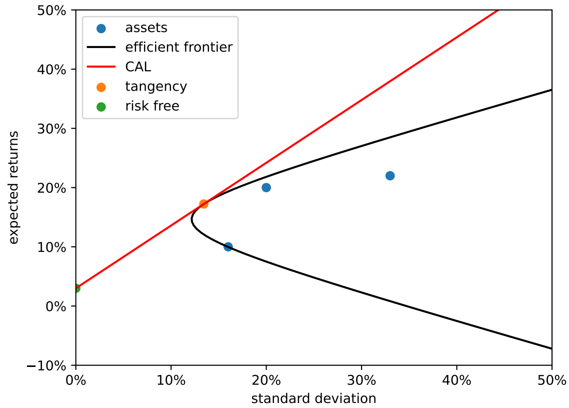


Figure 7: Capital Allocation Line (CAL) and Efficient Frontier

Proof. We write the weights of the first optimal portfolio as

$$\begin{aligned} w_\triangledown^{(1)} &= \frac{C - B\mu_P^{(1)}}{D}\Sigma^{-1}\mathbf{1} + \frac{A\mu_P^{(1)} - B}{D}\Sigma^{-1}\mu \\ &= \frac{C}{D}\Sigma^{-1}\mathbf{1} - \mu_P^{(1)}\frac{B}{D}\Sigma^{-1}\mathbf{1} + \mu_P^{(1)}\frac{A}{D}\Sigma^{-1}\mu - \frac{B}{D}\Sigma^{-1}\mu \\ &= \mu_P^{(1)}\underbrace{\left(\frac{A}{D}\Sigma^{-1}\mu - \frac{B}{D}\Sigma^{-1}\mathbf{1}\right)}_{=:u} + \underbrace{\frac{C}{D}\Sigma^{-1}\mathbf{1} - \frac{B}{D}\Sigma^{-1}\mu}_{=:v} \\ &= \mu_P^{(1)} u + v \end{aligned}$$

where $u, v \in \mathbb{R}^d$ are vectors, which do not depend on $\mu_P^{(1)}$. With the same argument

$$w_\triangledown^{(2)} = \mu_P^{(2)} u + v \qquad\qquad w_\triangledown^{(3)} = \mu_P^{(3)} u + v.$$

Therefore,

$$\begin{aligned} &\frac{\mu_P^{(3)} - \mu_P^{(2)}}{\mu_P^{(1)} - \mu_P^{(2)}} w_\triangledown^{(1)} + \frac{\mu_P^{(1)} - \mu_P^{(3)}}{\mu_P^{(1)} - \mu_P^{(2)}} w_\triangledown^{(2)} = \frac{\mu_P^{(3)} - \mu_P^{(2)}}{\mu_P^{(1)} - \mu_P^{(2)}}(\mu_P^{(1)} u + v) + \frac{\mu_P^{(1)} - \mu_P^{(3)}}{\mu_P^{(1)} - \mu_P^{(2)}}(\mu_P^{(2)} u + v) \\ &= \left(\frac{\mu_P^{(3)} - \mu_P^{(2)}}{\mu_P^{(1)} - \mu_P^{(2)}}\mu_P^{(1)} + \frac{\mu_P^{(1)} - \mu_P^{(3)}}{\mu_P^{(1)} - \mu_P^{(2)}}\mu_P^{(2)}\right) u + \left(\frac{\mu_P^{(3)} - \mu_P^{(2)}}{\mu_P^{(1)} - \mu_P^{(2)}} + \frac{\mu_P^{(1)} - \mu_P^{(3)}}{\mu_P^{(1)} - \mu_P^{(2)}}\right) v \\ &= \left(\frac{\mu_P^{(3)}\mu_P^{(1)} - \mu_P^{(2)}\mu_P^{(1)} + \mu_P^{(1)}\mu_P^{(2)} - \mu_P^{(3)}\mu_P^{(2)}}{\mu_P^{(1)} - \mu_P^{(2)}}\right) u + \left(\frac{\mu_P^{(3)} - \mu_P^{(2)} + \mu_P^{(1)} - \mu_P^{(3)}}{\mu_P^{(1)} - \mu_P^{(2)}}\right) v \\ &= \mu_P^{(3)} u + v = w_\triangledown^{(3)}. \end{aligned}$$

□

5. Capital Allocation Line (CAL)

In a market where the returns r of the risky assets are Gaussian, $r \sim \mathcal{N}(\mu, \Sigma)$, we have established the optimal portfolio given any target level of risk (or return). What more do we need?

Our derivation of the efficient frontier in (4.5) rests on the assumption that the covariance matrix Σ is positive-definite. While this might sound like a rather technical detail, this has a very concrete consequence: we are assuming that the all assets are risky. This means that when investing into this asset universe, Corollary 4.3 establishes a globally minimal amount of risk one has to be willing to take to make that investment.

This is problematic in two ways. First, if an investor is not comfortable with the risk in the minimum variance portfolio, then according to the theory established to far, one would have to walk away from that asset universe. This significantly limits the applicability of this theory. Second, the assumption that there is no risk-free asset is often deemed unrealistic.[5] Irrespective of whether or not we think truly risk-free assets exist, it would be highly desirable to be able to include a risk-free asset into our theory. From a theoretical point of view, one should remark that many other landmark theories in financial mathematics, for example Black-Scholes option pricing, do assume the existence of a risk-free asset. From a practical point of view, risky vs. risk-free is often seen as the difference between saving and investing. Many investors consider it unrealistic that a market has no risk-free asset due to the existence of e.g. bank account deposits.

Integrating a risk-free asset into the theory established so far is unfortunately not straightforward. By definition, a risk-free asset has zero volatility and this also means it has zero correlation with all the other assets. Thus, if only one of our d assets is risk-free, the resulting covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$ is no longer positive definite. While this violation of the assumptions of our theorems so far sounds like a technical detail, the consequence is that the inverse Σ^{-1} no longer exists. Given this expression appears in all the key equations, e.g. Eqs. (4.1), (4.2), (4.8) and (4.9), none of these expressions can be calculated anymore.

Even more importantly, the conclusions of these theorems, in particular (4.3), which states that the efficient frontier is a parabola, are all wrong. It turns out that introducing a risk-free asset changes the shape of the efficient frontier of the universe of $d+1$ assets, i.e. d risky assets and one risk-free asset, from a parabola to a straight line, which is called *Capital Allocation Line (CAL)*, see Fig. 7. In the following we derive this result and then explore the relationship between the efficient frontier of risky assets and the CAL. We will prove that they intersect tangentially in precisely one point. This *tangency portfolio* represents the entire market of risky assets and can be used to design portfolios with arbitrary risk-return profiles. While the intercept of the CAL is just the risk-free rate, its slope is called *Sharpe ratio*. It turns out that Sharpe ratios provide an alternative way to derive these results and are a very useful concept in general.

5.1. One Fund Theorem (Risk Minimization) In this section we prove that in presence of a risk-free asset, the efficient frontier takes the shape of a line.

Theorem 5.1 (One Fund Theorem, Risk Minimization)**.** Assume that the returns of d risky assets are Gaussian, $r \sim \mathcal{N}(\mu, \Sigma)$, and that in addition there exists one risk-free asset with expected return μ_f and zero variance. For any given target return $\bar{\mu}_P$, the problem

$$\text{minimize } w \mapsto \frac{1}{2} w^\top \Sigma w \quad \text{subject to } \sum_{i=1}^{d} w_i + w_f = 1 \quad \text{and} \quad \sum_{i=1}^{d} w_i \mu_i + w_f \mu_f = \bar{\mu}_P,$$

[5] At least in during Markowitz' times. Since the Great Financial Crisis in 07/08, working under the assumption that risk-free assets or risk-free rates do not exist in reality, has gained some more traction again.

has a unique solution

$$w_{\blacktriangledown} = w_{\blacktriangledown}(\bar{\mu}_P) = \frac{\bar{\mu}_P - \mu_f}{S^2}\Sigma^{-1}\mu_e, \tag{5.1}$$

where

$$\mu_e := \mu - \mu_f\,\mathbf{1}, \qquad S^2 := \mu_e^\top \Sigma^{-1}\mu_e \tag{5.2}$$

The resulting minimal variance is given by

$$(\sigma_P^{\blacktriangledown})^2 = \sigma_P^{\blacktriangledown}(\bar{\mu}_P)^2 = \frac{(\bar{\mu}_P - \mu_f)^2}{S^2}. \tag{5.3}$$

◊

Proof. Notice that while the problem is formulated in terms of $d+1$ weights, i.e. the d weights w of the risky assets and the one weight w_f for the risk-free asset. This first constraint can compactly be re-written as

$$w_f = 1 - \mathbf{1}^\top w$$

and plugged into the second we obtain

$$\bar{\mu}_P = \mu^\top w + (1 - \mathbf{1}^\top w)\mu_f = \mu_f + w^\top(\mu - \mu_f\,\mathbf{1}) = \mu_f + w^\top \mu_e.$$

Thus, we can still treat this as a d-dimensional optimization problem, drop first constraint and replace the second one with the above equation. The resulting Euler Lagrange equation from Theorem 8.1 gives

$$\Sigma w_{\blacktriangledown} = \lambda\mu_e \qquad \Longrightarrow w_{\blacktriangledown} = \lambda\Sigma^{-1}\mu_e.$$

We obtain

$$\bar{\mu}_P - \mu_f = w_{\blacktriangledown}^\top \mu_e = \lambda\mu_e^\top\Sigma^{-1}\mu_e = \lambda S^2$$

and therefore

$$\lambda = \frac{\bar{\mu}_P - \mu_f}{S^2}.$$

This implies

$$(\sigma_P^{\blacktriangledown})^2 = w_{\blacktriangledown}^\top \Sigma w_{\blacktriangledown} = \lambda^2\mu_e^\top\Sigma^{-1}\mu_e = \frac{(\bar{\mu}_P - \mu_f)^2}{S^4}S^2 = \frac{(\bar{\mu}_P - \mu_f)^2}{S^2}.$$

□

This theorem shows us that the means by which we can now achieve arbitrary low risk is the factor $\bar{\mu}_P - \mu_f$, which is called *excess returns*. Technically, this terminology only makes sense when $\bar{\mu}_P > \mu_f$. On the other hand, financially it doesn't make any sense to target a return that is less or equal than the risk-free rate using risky assets.

We can now easily derive the efficient frontier of the asset universe that includes the risk-free asset.

Corollary 5.2. The target returns can be written as a function of the minimal variance by

$$\bar{\mu}_P = \mu_f + \sigma_P^{\blacktriangledown} S. \tag{5.4}$$

This function is called the *Capital Allocation Line (CAL)*. ◊

Proof. We calculate

$$\begin{aligned}\bar{\mu}_P = \mu_f + w_{\blacktriangledown}^\top \mu_e &= \mu_f + (\bar{\mu}_P - \mu_f)\frac{\mu_e^\top \Sigma^{-1}\mu_e}{\mu_e^\top \Sigma^{-1}\mu_e} \\ &= \mu_f + \frac{\bar{\mu}_P - \mu_f}{S}\frac{S^2}{S} = \mu_f + \sigma_P^{\blacktriangledown} S\end{aligned}$$

□

5.2. One Fund Theorem (Return Maximization) Just like for the risky assets, one can derive the CAL either by minimizing variance given the target returns or by maximizing returns given the target variance. We have already given the former and even this is a bit redundant, we also give the later derivation here.

Theorem 5.3 (One Fund Theorem, Risk Minimization)**.** Assume that the returns of d risky assets are Gaussian, $r \sim \mathcal{N}(\mu, \Sigma)$, and that in addition there exists one risk-free asset with expected return μ_f and zero variance. For any given target variance $\bar{\sigma}_P^2$, the problem

$$\text{maximize } (w, w_f) \mapsto \sum_{i=1}^d w_i\mu_i + w_f\mu_f \quad \text{subject to } \sum_{i=1}^d w_i + w_f = 1 \quad \text{and} \quad w^\top \Sigma w = \sigma_P^2$$

has a unique solution

$$w_{\blacktriangle} = w_{\blacktriangle}(\bar{\sigma}_P) = \frac{\sigma_P}{S}\Sigma^{-1}\mu_e. \tag{5.5}$$

The resulting maximal return is given by

$$\mu_P^{\blacktriangle} = \mu_P^{\blacktriangle}(\sigma_P) = \sigma_P S + \mu_f. \tag{5.6}$$

◊

Proof. We again use the first constraint to write

$$w_f = 1 - \mathbf{1}^\top w.$$

Plugging this in, we have to maximize

$$w \mapsto \mu^\top w + (1 - \mathbf{1}^\top w)\mu_f = \mu_e^\top w + \mu_f.$$

Invoking Theorem 8.1 on this function and the equivalent second constraint $\frac{1}{2}w^\top \Sigma w = \frac{1}{2}\sigma_P^2$ (dropping the first constraint) we obtain

$$\mu_e = \lambda \Sigma w_{\blacktriangle}.$$

Again, if $\lambda = 0$, we had $\mu_e = 0$ and μ and $\mathbf{1}$ were linearly dependent contradicting the assumptions Assumption 3.1. Thus, $\lambda \neq 0$ and we obtain

$$w_{\blacktriangle} = \frac{1}{\lambda}\Sigma^{-1}\mu_e.$$

Plugging this into the constraint yields

$$\sigma_P^2 = w_{\blacktriangle}^\top \Sigma w_{\blacktriangle} = \frac{1}{\lambda^2}\mu_e^\top \Sigma^{-1}\mu_e = \frac{S^2}{\lambda^2} \Longrightarrow \lambda = \frac{S}{\sigma_P}.$$

Consequently,

$$w_{\blacktriangle} = \frac{\sigma_P}{S}\Sigma^{-1}\mu_e$$

This implies

$$\mu_P^{\blacktriangle} = \frac{\sigma_P}{S}\mu_e^\top \Sigma^{-1}\mu_e + \mu_f = \sigma_P S + \mu_f.$$

□

Remark 5.4. Just like in the risky case, these two derivations are equivalent. We can directly see from Eqs. (5.3) and (5.6) that these are consistent. Analogous to Theorem 4.7, we again also have for any target returns $\bar{\mu}_P$

$$w_{\blacktriangle}(\sigma_P^{\blacktriangledown}(\bar{\mu}_P)) \overset{(5.5)}{=} \frac{\sigma_P^{\blacktriangledown}}{S}\Sigma^{-1}\mu_e = \frac{\frac{\bar{\mu}_P-\mu_f}{S}}{S}\Sigma^{-1}\mu_e = \frac{\bar{\mu}_P-\mu_f}{S^2}\Sigma^{-1}\mu_e \overset{(5.1)}{=} w_{\blacktriangledown}(\bar{\mu}_P)$$

and for any target variance $\bar{\sigma}_P$

$$w_{\blacktriangledown}(\mu_P^{\blacktriangle}(\bar{\sigma}_P)) \overset{(5.1)}{=} \frac{\mu_P^{\blacktriangle}(\bar{\sigma}_P)-\mu_f}{S^2}\Sigma^{-1}\mu_e = \frac{\bar{\sigma}_P S}{S^2}\Sigma^{-1}\mu_e = \frac{\bar{\sigma}_P}{S}\Sigma^{-1}\mu_e \overset{(5.5)}{=} w_{\blacktriangle}(\bar{\sigma}_P). \qquad \diamondsuit$$

5.3. Tangency Portfolio In the example Fig. 7, we can see that in presence of a risk-free asset the resulting CAL touches the efficient frontier tangentially in precisely one point. This is not a coincidence, which we prove in the following.

Theorem 5.5 (Tangency Portfolio)**.** Assume that the returns of d risky assets are Gaussian, $r \sim \mathcal{N}(\mu, \Sigma)$, and that in addition there is one risk-free asset with expected return $\mu_f \le \mu_{\text{mvp}}$, c.f.(4.6). Then the efficient frontier

$$\sigma_P \mapsto \mu_P(\sigma_P) = \frac{\sqrt{D}}{A}\sqrt{A\sigma_P^2 - 1} + \frac{B}{A}$$

of the risky assets, c.f. (4.7), intersects the CAL, c.f. (5.4),

$$\sigma_P \mapsto \bar{\mu}_P(\sigma_P) = \mu_f + \sigma_P S$$

tangentially in precisely one point (μ_T, σ_T). This point is given by

$$\sigma_T = \frac{\sqrt{\mu_e^\top \Sigma^{-1}\mu_e}}{\mathbf{1}^\top \Sigma^{-1}\mu_e}, \qquad \mu_T = \frac{\mu^\top \Sigma^{-1}\mu_e}{\mathbf{1}^\top \Sigma^{-1}\mu_e}. \tag{5.7}$$

The portfolio underlying this point is called the *tangency portfolio* and its asset weights satisfy

$$w_T = \frac{\Sigma^{-1}\mu_e}{\mathbf{1}^\top \Sigma^{-1}\mu_e}, \qquad w_{T,f} = 0, \tag{5.8}$$

i.e. it is composed of risky assets only. $\diamondsuit$

Proof. The overall strategy is to show that the slopes of both equations, i.e. the CAL and the efficient frontier are equal in only one point and then we show that both functions have the same value there. To ease calculations, we first establish some algebraic relations.

STEP 1: When we derived the efficient frontier of the risky assets, we have introduced the quantities A, B, C, D in (4.2), which are convenient to work with on the parabola. For the efficient frontier in presence of a risk-free asset, we have introduced the quantities μ_e and S in (5.2) as these are convenient when working with that line. It is helpful to establish some relations between these quantities. First of all, notice that

$$\begin{aligned} S^2 &= \mu_e^\top \Sigma^{-1}\mu_e = (\mu - \mu_f\,\mathbf{1})^\top \Sigma^{-1}(\mu - \mu_f\,\mathbf{1}) \\ &= \mu^\top \Sigma^{-1}\mu - 2\mu_f \mu^\top \Sigma^{-1}\,\mathbf{1} + \mu_f^2\,\mathbf{1}^\top \Sigma^{-1}\,\mathbf{1} \\ &= C - 2\mu_f B + \mu_f^2 A \end{aligned}$$

and therefore

$$\begin{aligned}(B - A\mu_f)^2 + D &= B^2 - 2AB\mu_f + \mu_f^2 A^2 + (AC - B^2)\\ &= AC - 2AB\mu_f + A^2\mu_f^2 = AS^2.\end{aligned} \tag{5.9}$$

By hypothesis we have

$$\mu_f \le \mu_{\text{mvp}} \Longrightarrow B - A\mu_f \ge B - A\mu_{\text{mvp}} \overset{(4.6)}{=} B - A\frac{B}{A} = 0$$

and therefore

$$B - A\mu_f = \sqrt{AS^2 - D}.$$

Furthermore, we calculate

$$\mathbf{1}^\top \Sigma^{-1}\mu_e = \mathbf{1}^\top \Sigma^{-1}(\mu - \mu_f\,\mathbf{1}) = \mathbf{1}^\top \Sigma^{-1}\mu - \mu_f\,\mathbf{1}^\top \Sigma^{-1}\,\mathbf{1} = B - \mu_f A.$$

Combining these equations, we obtain

$$\mathbf{1}^\top \Sigma^{-1}\mu_e = B - \mu_f A = \sqrt{AS^2 - D}. \tag{5.10}$$

We also calculate

$$\mu^\top \Sigma^{-1}\mu_e = \mu^\top \Sigma^{-1}(\mu - \mu_f\,\mathbf{1}) = \mu^\top \Sigma^{-1}\mu - \mu_f \mu^\top \Sigma^{-1}\,\mathbf{1} = C - \mu_f B. \tag{5.11}$$

Step 2 (Equate Slopes): We first show that both functions have the same slope in precisely one point: The slope of the CAL is clearly

$$\frac{\partial \bar{\mu}_P}{\partial \sigma_P} = S.$$

The slope of the parabola is

$$\frac{\partial \mu_P}{\partial \sigma_P} = \frac{\sqrt{D}}{A}\frac{2A\sigma_P}{2\sqrt{A\sigma_P^2 - 1}} = \frac{\sqrt{D}\sigma_P}{\sqrt{A\sigma_P^2 - 1}}.$$

These are equal precisely when

$$\begin{aligned}S = \frac{\sqrt{D}\sigma_P}{\sqrt{A\sigma_P^2 - 1}} &\Longleftrightarrow S^2 = \frac{D\sigma_P^2}{A\sigma_P^2 - 1} \Longleftrightarrow D\sigma_P^2 = S^2(A\sigma_P^2 - 1) = AS^2\sigma_P^2 - S^2\\ &\Longleftrightarrow S^2 = (AS^2 - D)\sigma_P^2 \Longleftrightarrow \sigma_P^2 = \frac{S^2}{AS^2 - D} =: \sigma_T^2.\end{aligned}$$

Step 3 (Plug into functions): The corresponding return on the risky frontier (i.e. the parabola) is

$$\begin{aligned}\mu_T &:= \frac{\sqrt{D}}{A}\sqrt{A\frac{S^2}{AS^2 - D} - 1} + \frac{B}{A} = \frac{\sqrt{D}}{A}\sqrt{\frac{AS^2 - AS^2 + D}{AS^2 - D}} + \frac{B}{A}\\ &= \frac{D}{A\sqrt{AS^2 - D}} + \frac{B}{A} = \frac{D + B\sqrt{AS^2 - D}}{A\sqrt{AS^2 - D}}.\end{aligned}$$

The corresponding return on the CAL (i.e. the line) is

$$\bar{\mu}_T := \mu_f + \sigma_T S = \mu_f + S\sqrt{\frac{S^2}{AS^2 - D}} = \frac{A\mu_f\sqrt{AS^2 - D} + AS^2}{A\sqrt{AS^2 - D}}$$

Step 4: To see that these points are equal, we first need to reformulate these quantities a bit using our preparations. We re-write the first via

$$\mu_T = \frac{D + B\sqrt{AS^2 - D}}{A\sqrt{AS^2 - D}} \overset{(5.10)}{=} \frac{D + B(B - A\mu_f)}{A\sqrt{AS^2 - D}} = \frac{D + B^2 - AB\mu_f}{A\sqrt{AS^2 - D}} = \frac{C - B\mu_f}{\sqrt{AS^2 - D}}.$$

We re-write the second as well via

$$\begin{aligned}
\bar{\mu}_T &\overset{(5.10)}{=} \frac{A\mu_f(B - A\mu_f) + AS^2}{A\sqrt{AS^2 - D}} \\
&\overset{(5.9)}{=} \frac{A\mu_f(B - A\mu_f) + (B - A\mu_f)^2 + D}{A\sqrt{AS^2 - D}} \\
&= \frac{AB\mu_f - A^2\mu_f^2 + B^2 - 2AB\mu_f + A^2\mu_f^2 + AC - B^2}{A\sqrt{AS^2 - D}} \\
&= \frac{C - B\mu_f}{\sqrt{AS^2 - D}} = \mu_T
\end{aligned}$$

Thus, we have proven that the CAL intersect the efficient frontier of the risky assets tangentially in one and only one point.
Step 5: We can write the resulting quantities a bit nicer by calculating further

$$\mu_T = \frac{C - B\mu_f}{\sqrt{AS^2 - D}} \overset{(5.10)}{=} \frac{C - B\mu_f}{\mathbf{1}^\top \Sigma^{-1}\mu_e} \overset{(5.11)}{=} \frac{\mu^\top \Sigma^{-1}\mu_e}{\mathbf{1}^\top \Sigma^{-1}\mu_e}.$$

The volatility can also be expressed via

$$\sigma_T = \frac{S}{\sqrt{AS^2 - D}} = \frac{\sqrt{\mu_e^\top \Sigma^{-1}\mu_e}}{\mathbf{1}^\top \Sigma^{-1}\mu_e}.$$

Step 6: To see the claim about the weights we use (5.1) to obtain

$$\begin{aligned}
w_T &= (\bar{\mu}_T - \mu_f)\frac{\Sigma^{-1}\mu_e}{\mu_e^\top \Sigma^{-1}\mu_e} = \left(\frac{\mu^\top \Sigma^{-1}\mu_e}{\mathbf{1}^\top \Sigma^{-1}\mu_e} - \mu_f\right)\frac{\Sigma^{-1}\mu_e}{\mu_e^\top \Sigma^{-1}\mu_e} \\
&= \frac{\mu^\top \Sigma^{-1}\mu_e - \mu_f \mathbf{1}^\top \Sigma^{-1}\mu_e}{\mathbf{1}^\top \Sigma^{-1}\mu_e}\frac{\Sigma^{-1}\mu_e}{\mu_e^\top \Sigma^{-1}\mu_e} = \frac{\mu^\top - \mu_f \mathbf{1}^\top}{\mathbf{1}^\top \Sigma^{-1}\mu_e}\Sigma^{-1}\mu_e\frac{\Sigma^{-1}\mu_e}{\mu_e^\top \Sigma^{-1}\mu_e} \\
&= \frac{\mu_e^\top}{\mathbf{1}^\top \Sigma^{-1}\mu_e}\Sigma^{-1}\mu_e\frac{\Sigma^{-1}\mu_e}{\mu_e^\top \Sigma^{-1}\mu_e} = \frac{\mu_e^\top \Sigma^{-1}\mu_e}{\mathbf{1}^\top \Sigma^{-1}\mu_e}\frac{\Sigma^{-1}\mu_e}{\mu_e^\top \Sigma^{-1}\mu_e} \\
&= \frac{\Sigma^{-1}\mu_e}{\mathbf{1}^\top \Sigma^{-1}\mu_e}
\end{aligned}$$

Since clearly,

$$\mathbf{1}^\top w_T = \frac{1}{\mathbf{1}^\top \Sigma^{-1}\mu_e}\mathbf{1}^\top \Sigma^{-1}\mu_e = 1,$$

these weights already sum to 1, thus no fraction is in the risk-free part, i.e. $w_{T,f} = 0$. □

5.4. Sharpe Ratio The slope S of the CAL has an alternative expression other than (5.2) and a famous name.

Lemma 5.6. The quantity S satisfies

$$S = \frac{\mu_T - \mu_f}{\sigma_T} \tag{5.12}$$

and is also called the *Sharpe Ratio* of the tangency portfolio. ◊

Proof. We calculate

$$\frac{\mu_T - \mu_f}{\sigma_T} = \frac{\frac{\mu^\top \Sigma^{-1}\mu_e}{\mathbf{1}^\top \Sigma^{-1}\mu_e} - \mu_f}{\frac{\sqrt{\mu_e^\top \Sigma^{-1}\mu_e}}{\mathbf{1}^\top \Sigma^{-1}\mu_e}} = \frac{\mathbf{1}^\top \Sigma^{-1}\mu_e}{\sqrt{\mu_e^\top \Sigma^{-1}\mu_e}} \frac{\mu^\top \Sigma^{-1}\mu_e - \mu_f \mathbf{1}^\top \Sigma^{-1}\mu_e}{\mathbf{1}^\top \Sigma^{-1}\mu_e}$$
$$= \frac{\mu^\top \Sigma^{-1}\mu_e - \mu_f \mathbf{1}^\top \Sigma^{-1}\mu_e}{\sqrt{\mu_e^\top \Sigma^{-1}\mu_e}} = \frac{\mu_e^\top \Sigma^{-1}\mu_e}{\sqrt{\mu_e^\top \Sigma^{-1}\mu_e}} = S. \qquad \square$$

The Sharpe ratio is a very interesting quantity. It represents the excess returns per unit of risk taken. Many investors find it natural to maximize this quantity as it provides an aggregate view of jointly maximizing excess return (nominator) and minimizing risk (denominator).

5.5. Two Fund Separation Theorem Theorem 5.5 is a remarkable result. It shows that the CAL has two distinct points that define it: The point $(0, \mu_f)$, where everything is invested in the risk-free asset, and (σ_T, μ_T) where everything is invested in risky assets. Recall from Theorem 4.8 that even without the risk-free asset, it is possible to produce any point on the efficient frontier of the risky assets from just two distinct points on it. But with the risky assets only, there are no two distinct points on the efficient frontier to use. With the risk-free asset though, we can now produce any point on the CAL by using just these two canonical portfolios behind these two distinct points. This is called the two fund separation theorem.

Theorem 5.7 (two fund separation theorem)**.** For any target volatility σ_P, the portfolio that achieves optimal returns $\bar{\mu}_P(\sigma_P)$, i.e. such that $(\sigma_P, \bar{\mu}_P))$ lie on the CAL can be constructed by purchasing the amounts

$$w_f(\sigma_P) := \frac{\sigma_T - \sigma_P}{\sigma_T}, \qquad w_T(\sigma_P) := \frac{\sigma_P}{\sigma_T} \tag{5.13}$$

of the risk-free asset, respectively, of the tangency portfolio. ◊

Proof. We simply use the CAL formula (5.4), plug in (5.12) and rearrange to

$$\bar{\mu}_P(\sigma_P) = \mu_f + \sigma_P S = \mu_f + \sigma_P \frac{\mu_T - \mu_f}{\sigma_T}$$
$$= \mu_f + \sigma_P \frac{\mu_T}{\sigma_T} - \sigma_P \frac{\mu_f}{\sigma_T} = \frac{\sigma_T - \sigma_P}{\sigma_T}\mu_f + \frac{\sigma_P}{\sigma_T}\mu_T. \qquad \square$$

5.6. Sharpe Ratio Maximization Sharpe ratios can be calculated of any asset or portfolio, not just the tangency portfolio. A practical approach to portfolio management can be to maximize the Sharpe ratio. It turns out that this approach paves an alternative way to introduce the tangency portfolio.

Definition 5.8. Let $\bar{\mu}_P$ be the returns of a portfolio P with weight $w_f \in \mathbb{R}$ in the risk-free asset and weights $w \in \mathbb{R}^d$ in the assets. Let σ_P be the volatility of the resulting portfolio. Then

$$S_P := S(w_f, w) := \frac{\bar{\mu}_P - \mu_f}{\sigma_P}$$

is called the *Sharpe ratio* of P. ◊

Because the weights have to sum to one, there is an easier representation of this quantity.

Lemma 5.9 (representation of the Sharpe ratio)**.** The Sharpe ratio S_P of any portfolio P can be written as a function of the weights $w \in \mathbb{R}^d$ of the risky assets as

$$S_P = S(w) = \frac{\mu_e^\top w}{\sqrt{w^\top \Sigma w}}. \tag{5.14}$$

$\diamondsuit$

Proof. Recall that for any choice of weight $w_f \in \mathbb{R}$ for the risk-free asset and any choice of weights $w \in \mathbb{R}^d$ for the risky assets, the return on the portfolio as a function of the weights is given by

$$\bar{\mu}_P = \mu_P(w_f, w) = w_f \mu_f + \mu^\top w,$$

but this is subject to the constraint

$$1 = w_f + \mathbf{1}^\top w \Longrightarrow w_f = 1 - \mathbf{1}^\top w.$$

Consequently, we can also write the portfolio returns as a function of the risky weights only as

$$\mu_P = \mu_P(w) = (1 - \mathbf{1}^\top w)\mu_f + \mu^\top w = \underbrace{(\mu - \mu_f \mathbf{1})^\top}_{=\mu_e} w + \mu_f.$$

Because the risk-free weight has zero variance, we can write the Sharpe ratio S of the portfolio as a function of the risky weights w only as

$$S(w) = \frac{\mu_e^\top w}{\sqrt{w^\top \Sigma w}}. \qquad \square$$

Theorem 5.10. Let $r \sim \mathcal{N}(\mu, \Sigma)$ and μ_f be a risk-free rate. All solutions that solve the unconstrained optimization problem

$$\text{maximize } w \mapsto S(w) := \frac{w^\top \mu_e}{\sqrt{w^\top \Sigma w}},$$

are given by

$$\{\kappa w_T \mid \kappa \in \mathbb{R}\},$$

where w_T is the tangency portfolio. The volatilities and returns of the portfolios corresponding to these risky assets are exactly the CAL. $\diamondsuit$

Proof. Notice that the Sharpe ratio function (5.14) is invariant under scalar multiplication, i.e. for any $\kappa \in \mathbb{R}$ and any $w \in \mathbb{R}^d$,

$$S(\kappa w) = S(w).$$

Thus if we can find one maximum, we just have to scale it to find all. We therefore can equivalently solve the constrained problem

$$\text{maximize } w \mapsto \mu_e^\top w \text{ subject to } w^\top \Sigma w = 1.$$

Again by Theorem 8.1, the optimal value $w_\diamond$ satisfies

$$\mu_e^\top = \lambda \frac{1}{2} w_\diamond^\top \Sigma.$$

Notice that if $\lambda = 0$, this implies $\mu_e = 0$ and hence μ and $\mathbf{1}$ were linearly dependent, which contradicts our assumptions Assumption 3.1, hence we obtain $\lambda \neq 0$. This implies

$$\mu_e = \frac{\lambda}{2}\Sigma w_\diamond \implies \frac{2}{\lambda}\Sigma^{-1}\mu_e = w_\diamond.$$

Plugging this into the constraint gives

$$1 = w_\diamond^\top \Sigma w_\diamond = \frac{4}{\lambda^2}\mu_e^\top \Sigma^{-1}\mu_e \implies \lambda = 2\sqrt{\mu_e^\top \Sigma^{-1}\mu_e}.$$

Thus, this optimal weight satisfies[6]

$$w_\diamond = \frac{2}{2\sqrt{\mu_e^\top \Sigma^{-1}\mu_e}}\Sigma^{-1}\mu_e = \frac{\Sigma^{-1}\mu_e}{\sqrt{\mu_e^\top \Sigma^{-1}\mu_e}} \overset{(5.8)}{=} \frac{\sqrt{\mu_e^\top \Sigma^{-1}\mu_e}}{\mathbf{1}^\top \Sigma^{-1}\mu_e} w_T,$$

i.e. it is a constant multiple of the tangency portfolio w_T. Therefore, this parametrizes the CAL by (5.13). □

6. Capital Asset Pricing Model (CAPM)

The Capital Asset Pricing Model is an idea going back to [5], which answers a seemingly simple question: how do the returns of an asset relate to the returns of a market? It turns out that if we accept Assumption 3.1 of the Gaussian model and its implications, then the answer is very simple: In presence of a risk-free asset, the excess returns $X_i := r_i - \mu_f$ of every asset i are a constant multiple $\beta_i \in \mathbb{R}$ of the excess returns $X_m := r_m - \mu_f$ of the market plus some non-systematic noise ε_i that is uncorrelated with the market. In particular, every assets expected long term return above the market is $\alpha_i = 0$.

In theory, the implications of this insight are that this model confirms the passive approach to investing. After all, if no asset yields an advantage what point is there in putting effort into security selection? One should simply buy one diversified market portfolio and wait.

In practice it is not that simple. Even in theory the conclusions of that model only strictly hold for choosing the tangency portfolio to represent the market, which in practice is often not the case. The conclusions rely on the Assumption 3.1 and it is known that in practice markets do not exactly follow a Gaussian distribution. For that reason another pragmatic use of the CAPM is to use it to derive benchmarks for active asset management.

6.1. Assets and Portfolios In order to be able to apply the CAPM to that wider use case, we we need a slightly more general version of it, which we derive now.

The following is an application of a stochastic version of the following simple geometric fact: Every vector X_i is a multiple β_i of every other vector X_P plus some residual ε_i that is orthogonal to X_P.

[6] Notice that since the volatility at that point is 1 by construction, this gives a Sharpe ratio of

$$S(w_\diamond) = \mu_e^\top w_\diamond = \frac{\mu_e^\top \Sigma^{-1}\mu_e}{\sqrt{\mu_e^\top \Sigma^{-1}\mu_e}} = \sqrt{\mu_e^\top \Sigma^{-1}\mu_e} \overset{(5.2)}{=} S.$$

Lemma 6.1. Let $r \sim \mathcal{N}(\mu, \Sigma)$ be the return vector of d risky assets, μ_f be a risk-free return and $r_P := w_P^\top r$ be the return of any portfolio with weights $w_P \in \mathbb{R}^d$. Then the excess returns $X_P := r_P - \mu_f$ of the portfolio and the excess returns $X_i := r_i - \mu_f$ of any asset satisfy

$$X_i = \beta_i X_P + A_i, \qquad\qquad A_i = \alpha_i + \varepsilon_i, \tag{6.1}$$

for some random variables A_i, ε_i and $\alpha_i \in \mathbb{R}$. These satisfy

$$\mathrm{Cov}[X_P, A_i] = \mathrm{Cov}[X_P, \varepsilon_i] = 0, \qquad\qquad \mathbb{E}[\varepsilon_i] = 0, \tag{6.2}$$

and

$$\beta_i = \frac{\mathrm{Cov}[X_i, X_P]}{\mathbb{V}[X_P]}, \qquad\qquad \alpha_i = \mathbb{E}[A_i] = \mu_i - \mu_f - \beta_i(\mu_P - \mu_f). \tag{6.3}$$

In particular, their conditional expectation is given by

$$\mathbb{E}[X_i \mid X_P] = \alpha_i + \beta_i X_P. \tag{6.4}$$

◊

Proof. Consider the Hilbert space of L^2 random variables and define the 2-dimensional linear subspace $U := \mathrm{span}(1, X_P)$. By definition this means, there exist unique $\alpha_i, \beta_i \in \mathbb{R}$ and $\varepsilon_i \in L^2$ such that

$$X_i = (\alpha_i + \beta_i X_P) + \varepsilon_i \in U \oplus V,$$

where V is the orthogonal complement to U. In particular, setting $A_i := \alpha_i + \varepsilon_i$, this proves (6.1). By definition of orthogonality, we obtain

$$\begin{aligned} 0 &= \langle \varepsilon_i, 1 \rangle_{L^2} = \mathbb{E}[\varepsilon_i], \\ 0 &= \langle \varepsilon_i, X_P \rangle_{L^2} = \mathbb{E}[\varepsilon_i X_P], \end{aligned}$$

thus

$$\begin{aligned} &\mathrm{Cov}[\varepsilon_i, X_P] = \mathbb{E}[\varepsilon_i X_P] - \mathbb{E}[\varepsilon_i]\,\mathbb{E}[X_P] = 0, \\ \Longrightarrow\ &\mathrm{Cov}[X_P, A_i] = \mathrm{Cov}[X_P, \varepsilon_i] = 0, \end{aligned}$$

since α_i is a constant. This shows (6.2) and implies

$$\mathrm{Cov}[X_i, X_P] = \beta_i\, \mathbb{V}[X_P],$$

as well as

$$\alpha_i = \mathbb{E}[A_i] = \mathbb{E}[X_i] - \beta_i\,\mathbb{E}[X_P] = \mu_i - \mu_f - \beta_i(\mu_P - \mu_f).$$

hence we have shown (6.4). To see the claim on conditional expectations, recall that since the d-vector of assets r is Gaussian, the variable (X_i, X_P) is jointly Gaussian as well. Hence, the conditional expectation is given by[7]

$$\begin{aligned} \mathbb{E}[X_i \mid X_P] &= \mathbb{E}[X_i] + \frac{\mathrm{Cov}[X_i, X_P]}{\mathbb{V}[X_P]}(X_P - \mathbb{E}[X_P]) \\ &= \mu_i - \mu_f + \beta_i(X_P - (\mu_P - \mu_f)) \\ &= \mu_i - \mu_f - \beta_i(\mu_P - \mu_f) + \beta_i X_P \\ &= \alpha_i + \beta_i X_P \end{aligned}$$

□

[7] recall e.g. https://en.wikipedia.org/wiki/Multivariate_normal_distribution#Conditional_distributions

6.2. CAPM Formulations Armed with Lemma 6.1 we can now prove that if we choose the tangency portfolio as the portfolio P, we obtain the equation that made the CAPM famous.

Theorem 6.2. Let $r \sim \mathcal{N}(\mu, \Sigma)$ be the d-dimensional random variable of the risky asset returns and $r_T := w_T^\top r$ be the returns of the tangency portfolio. Then the relationship between the excess returns of asset i and the excess returns of the tangency portfolio are given by

$$\mu_i - \mu_f = \beta_i(\mu_T - \mu_f), \qquad \beta_i := \frac{\mathrm{Cov}[r_i, r_T]}{\mathbb{V}[r_T]}. \tag{6.5}$$

In particular,

$$\alpha_i = \mathbb{E}[A_i] = 0. \tag{6.6}$$

$\Diamond$

Proof. The covariance matrix between the return vector r and the portfolio returns r_T is given by

$$\mathrm{Cov}[r, r_T] = \mathrm{Cov}[r, r]w_T = \Sigma w_T.$$

Using (5.1) for $\bar{\mu}_P = \mu_T$ we obtain

$$w_T = \frac{\mu_T - \mu_f}{S^2}\Sigma^{-1}\mu_e \iff \mu_e = \frac{S^2}{\mu_T - \mu_f}\Sigma w_T = \frac{S^2}{\mu_T - \mu_f}\mathrm{Cov}[r, r_T]$$

Plugging the definition of μ_e we obtain

$$\begin{aligned}\mu - \mu_f\,\mathbf{1} &= \mathrm{Cov}[r, r_T]\frac{S^2}{\mu_T - \mu_f} = \mathrm{Cov}[r, r_T]\frac{S^2}{(\mu_T - \mu_f)^2}(\mu_T - \mu_f) \\ &\overset{(5.3)}{=} \frac{\mathrm{Cov}[r, r_T]}{\sigma_T^2}(\mu_T - \mu_f) = \frac{\mathrm{Cov}[r, r_T]}{\mathbb{V}[r_T]}(\mu_T - \mu_f).\end{aligned}$$

Evaluating this equation of vectors for any component i gives the result. $\square$

Remark 6.3 (market vs tangency portfolio)**.** In practice, the CAPM is often formulated using the term *market portfolio* r_m instead of tagency portfolio r_T. Even though strictly the results only hold for the tangency portfolio, some may find it useful to think about returns in this way. $\Diamond$

Remark 6.4 (CAPM formulations)**.** There are three slightly different formulations of the CAPM depending on whether we are interested in returns, conditional expectations or expectations. We use the notation r_m for the market portfolio here.

(i) **Asset Return Generation Model**

$$r_i = \mu_f + \beta_i(r_m - \mu_f) + \varepsilon_i \tag{6.7}$$

This is an equality of random variables obtained from Eqs. (6.1) and (6.6), where

- r_i are the returns of the asset
- r_m are the returns of the market portfolio

- μ_f is the risk-free rate,
- $\beta_i = \frac{\mathrm{Cov}[r_m, r_i]}{\mathbb{V}[r_i]} = \frac{\sigma_i \rho_{im}}{\sigma_m} \in \mathbb{R}$ is the beta
- $\sigma_i^2 := \mathbb{V}[r_i] = \beta_i^2 \sigma_m^2 + \sigma_{\epsilon_i}^2$ is the total variance of the asset
- $\sigma_m^2 := \mathbb{V}[r_m]$ is the variance of the market
- ε_i is the error term, $\mathrm{Cov}[\varepsilon_i, r_m] = 0$, $\mathbb{E}[\varepsilon_i] = 0$.
- $\sigma_{\epsilon_i}^2 := \mathbb{V}[\varepsilon_i] = \sigma_i^2 - \beta_i^2 \sigma_m^2$ is the non-systematic variance
- $\beta_i^2 \sigma_m^2$ is the systematic variance of the asset
- σ_i is the total risk, $\beta_i \sigma_m$ the systematic risk and σ_ε the non-systematic risk of the asset.

(ii) **Conditionally expected asset returns given market returns**:

$$\mathbb{E}[r_i \mid r_m] = \mu_f + \beta_i (r_m - \mu_f) \tag{6.8}$$

This is an equality of random variables obtained from Eqs. (6.4) and (6.6), where

- $\mathbb{E}[r_i \mid r_m]$ are the conditionally expected asset returns r_i given the market returns r_m.

(iii) **Expected asset returns**:

$$\mu_i = \mathbb{E}[r_i] = \mu_f + \beta_i (\mu_m - \mu_f) \tag{6.9}$$

This is an equality of numbers obtained from (6.5), where

- $\mu_i = \mathbb{E}[r_i]$ are the expected returns of asset i
- $\mu_m = \mathbb{E}[r_m]$ are the expected returns of the market portfolio ◊

6.3. The Beta & (Non)systematic Risk The quantities β_i and σ_{ε_i} from (6.7) have a very concrete interpretation. It means that the exccess returns of an asset are driven by only two sources:

(i) **Systematic risk**: The asset moves because the market moves (precisely by $\beta_i(r_m - \mu_f)$. Financially, this is driven by macro-economic effects and is not diversifiable.

(ii) **Nonsystematic Risk:** This is the move due to ε_i. This risk is idiosyncratic and hence diversifiable.

The beta is a factor that translates how the market moves affect the asset moves. Depending on its size and sign, it has very different effects, see also Fig. 8:

- $\beta > 1$: The asset moves in the same direction as the market, but stronger.
- $0 < \beta \leq 1$: The asset moves in the same direction as the market, but weaker.
- $\beta = 0$: The asset has no systematic risk at all.
- $-1 < \beta < 0$: The asset moves in the opposite direction as the market, but with smaller magnitude.
- $\beta < -1$: The asset moves in the opposite direction as the market but with larger magnitude.

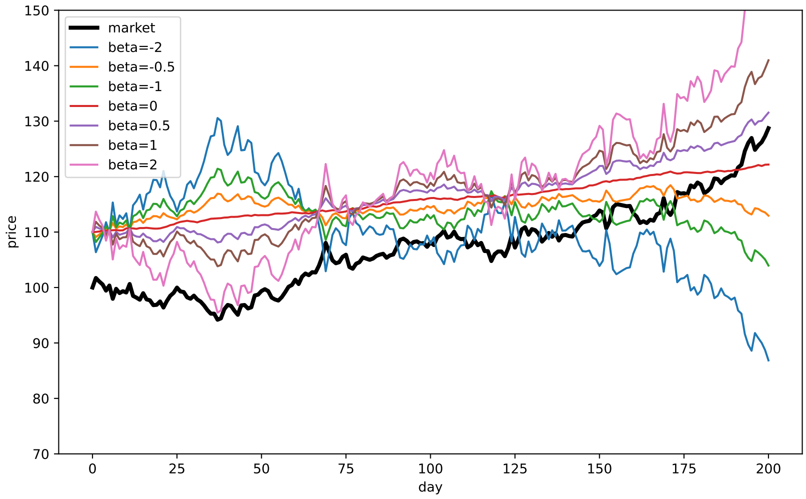


Figure 8: Effect of Beta

6.4. CAPM and Simple Regression One should remark that when we write down (6.1) again

$$X_i = \alpha_i + \beta_i X_m + \varepsilon_i \tag{6.10}$$

this equation bears close resemblance to *Simple Linear Regression*. In linear regression we assume that there is a pair of real-valued random variables (X, Y) and that these are theoretically related by

$$Y = f(X) + \varepsilon,$$

where ε satisfies $\mathbb{E}[\varepsilon \mid X] = 0$ and $f : \mathbb{R} \to \mathbb{R}$ is a function. Notice that this automatically implies $\mathbb{E}[Y \mid X] = f(X)$, i.e. f is the conditional expectation function. In practice, we are often given iid (x_j, y_j), $j = 1, \ldots, N$, of (X, Y), and hence condition this on $X = x_j$ and obtain

$$Y_j := f(x_j) + \varepsilon_j,$$

where $\varepsilon_j := \varepsilon \mid X = x_j$. In addition, one often makes the assumption that these residuals are Gaussian with constant variance, i.e. $\varepsilon_j \sim \mathcal{N}(0, \sigma^2)$. Such a regression is called *simple*[8], if we assume in addition that the relationship is linear, i.e.

$$f(x) = \beta x + \alpha,$$

where α is called *intercept* and β is called *slope*. Notice that in this case (6.10) is precisely an example of the simple linear regression setup where $Y = X_i$ and $X = X_m$.

Geometrically, the linear regression problem amounts to fitting a line to a cloud of points, see Fig. 9. In order to make precise what we mean by *fit*, one usually demands that the fitted line minimizes the squared error, i.e.

$$J(\alpha, \beta) := \sum_{i=j}^{N} (f(x_j) - y_j)^2.$$

[8] see e.g. https://en.wikipedia.org/wiki/Simple_linear_regression

Figure 9: CAPM and Linear Regression

One can show that the resulting coefficients $(\hat{\beta}, \hat{\alpha})$ that minimize this error are precisely given by

$$\hat{\beta} = \frac{\widehat{\mathrm{Cov}}[\hat{x}, \hat{y}]}{\sqrt{\widehat{\mathrm{Var}}[\hat{x}]}} = \frac{\sum_{j=1}^{N} (x_j - \bar{x}_j)(y_j - \bar{y}_j)}{\sum_{i=1}^{N} (x_j - \bar{x})^2}, \qquad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x},$$

where

$$\bar{x} := \frac{1}{N} \sum_{j=1}^{N} x_j, \qquad \bar{y} := \frac{1}{N} \sum_{j=1}^{N} y_j$$

are the estimated empirical means of X and Y and

$$\widehat{\mathrm{Cov}}[\hat{x}, \hat{y}] := \frac{1}{N} \sum_{j=1}^{N} (x_j - \bar{x})(y_j - \bar{y}), \qquad \widehat{\mathrm{Var}}[\hat{x}] := \frac{1}{N} \sum_{j=1}^{N} (x_j - \bar{x})^2$$

are the empirically estimated covariance between X and Y and the empirically estimated variance of X. Or in other words, this is exactly the empirical version of the theoretical result (6.4) for alpha and beta. For that reason, practitioners sometimes like to use the linear regression setup to estimate these quantities as linear regression has a rich statistical framework that allows for goodness of fit assessments and analysis.

6.5. Abnormal Returns The term $A_i = \alpha_i + \varepsilon_i$ from (6.1) is called *abnormal return*. This is a random variable to the sum of the constant α_i and the ε_i. The reason for that name is the following: Given a risk free rate μ_f, the β_i of an asset, we can work out the returns we expect on that asset i on any given day and given the r_m of the market under CAPM assumptions using (6.8):

$$\mathbb{E}[r_i \mid r_m] = \mu_f + \beta_i X_m = \mu_f + \beta_i (r_m - \mu_f)$$

This result will in general deviate from the returns r_i we actually observe on that day and the difference is precisely

$$r_i - \mathbb{E}[r_i \mid r_m] = r_i - \mu_f - \beta_i X_m = X_i - \beta_i X_m = \alpha_i + \varepsilon_i = A_i.$$

Unfortunately, the terminology in the literature is not entirely coherent. Sometimes the A_i is also called the (realized) α_i.

6.6. The Alpha & Performance Measurement Assume that P is a portfolio of assets with weights w. Then the excess returns of the portfolio satisfy

$$r_P - \mu_f = \sum_{i=1}^{d} w_i r_i - \mu_f \underbrace{\sum_{i=1}^{d} w_i}_{=1} = \sum_{i=1}^{d} w_i (r_i - \mu_f) = \sum_{i=1}^{d} w_i (\alpha_i + \beta_i (r_m - \mu_f) + \varepsilon_i)$$

and thus we can naturally assign an α_P, β_P and ε_P to the portfolio as well:

$$\alpha_P := \sum_{i=1}^{d} w_i \alpha_i, \qquad \beta_P := \sum_{i=1}^{d} w_i \beta_i, \qquad \varepsilon_P := \sum_{i=1}^{d} w_i \varepsilon_i.$$

One way to go about the question on whether or not it is a good idea to assume that Assumption 3.1 are correct, hence that putting effort into security selection is in vain, is to do it anyway and for any such portfolio P evaluate if it was worth doing. This is called *performance measurement* and the most obvious metric to measure performance is precisely that alpha:

Definition 6.5 (Jensen's alpha)**.** For any given portfolio P the metric

$$\alpha_P := r_P - \mu_f - \beta_P (r_m - \mu_f)$$

is called the *Jensen's alpha.* $\diamondsuit$

By MPT $\alpha_P = 0$, so over long periods of time, this metric should be zero. Hence, if one manages to assemble a portfolio where $\alpha_P > 0$ this hints at systematic risk-adjusted out-performance. The Jensen's alpha is precisely the expected abnormal return of the portfolio.

Notice though that there is a plethora of other performance metrics for asset managers.

Definition 6.6 (Information Ratio)**.** The information ratio is the abnormal return per unit of risk added by a security to a well-diversified portfolio

$$IR_P := \frac{\alpha_P}{\sigma_\varepsilon},$$

where σ_ε is the non-systematic risk. $\diamondsuit$

This can be used for security selection. The higher the IR, the more valuable the security.

The following is an extension of the Sharpe ratio.

Definition 6.7 (Treynor Ratio)**.** Let r_P be the returns of a portfolio with β_P over some risk-free asset with returns r_f. Then

$$T_P := \frac{r_P - r_f}{\beta_P}$$

is called the *Treynor ratio of P*. (This is only meaningful if $\beta_P > 0$). $\diamondsuit$

This can be used to rank individual securities.

There are many more such metrics and their usage comes with heavily debated pro's and cons.

7. Enduring Insights & Practical Limitations

Modern Portfolio Theory (MPT) and its subsequent adaptations CAL and CAPM can be considered breakthroughs of portfolio optimization theory of their time, the 1950's, which has been recognized through multiple Nobel prizes in the 1990's. Given this is now several decades in the past, we close with some remarks on how to think about this theory today. What has endured and what has not stood the test of time?

The insights that have endured most are the conceptual, big picture contributions, in particular:

(i) **Constrained Optimization Theory:** The realization that portfolio construction should be formulated as a risk-return trade-off, not merely a return maximization.

(ii) **Diversification:** The insight that the risk-return profile of a portfolio does not only depend on the risk-return profile of its composing assets, but materially on their interactions. This is the intellectual foundation of diversification.

(iii) **Geometry:** The geometric representation of optimal risk-return profiles leading to efficient frontiers.

The big strength of MPT is that these deep fundamental insights can be explored quantitatively in a very tractable analytic manner due to the assumption of jointly Gaussian asset returns.

However, that assumption is also at the heart of the model's greatest weaknesses:

(i) **Gaussianity:** Empirical analysis of asset returns show various behaviors that violate the Gaussian assumption. The probability of upside and downside moves is not symmetric due to skewness effects. Tail events representing large losses occur significantly more frequently than a Gaussian model would predict.

(ii) **Risk Metrics:** As a consequence, measuring portfolio risk using volatility as the only metric is not practical. However, replacing volatility by more advanced metrics such as VaR, ES or more general risk measures has led to interesting, but also more challenging risk-return optimization problems, see e.g. [1].

(iii) **Instability:** The key inputs needed for MPT are the returns, the volatilities and the correlations of the assets. Simply estimating these from historic data has proven difficult as the estimates subtly depend on observation windows and outliers reflecting that in reality the data is not only non-Gaussian, but also non-stationary. The efficient frontier is unfortunately quite sensitive to these estimates resulting in instabilities in optimal portfolio calculation.

8. Appendix

One of the most famous theorems in elementary calculus is that if a differentiable function $f : \mathbb{R} \to \mathbb{R}$ has an extremum at x^*, then

$$f'(x^*) = 0.$$

More generally, if a multi-variate differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ has an extremum at x^*, then

$$\nabla f(x^*) = 0.$$

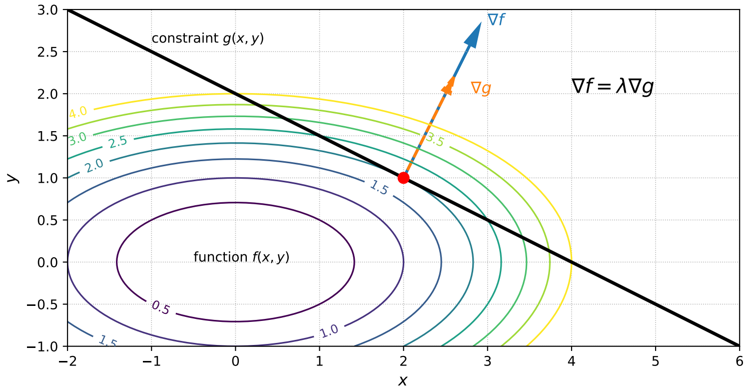


Figure 10: Lagrange Multiplier

Both of these cases deal with unconstrained optimality. When studying functions under constraints, the most famous case are differentiable equality constraints. The following theorem is well-known[9] and states that in this case the gradient of the objective f needs to be linearly dependent to the gradient of the constraints g, see Fig. 10.

Theorem 8.1 (Euler-Lagrange, [3, Sect 11.3])**.** Let $f : \mathbb{R}^d \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^k$ continuously differentiable functions. Assume that x^* is a solution of the constrained optimization problem

$$\text{minimize } f(x) \qquad\qquad \text{subject to } g(x) = 0.$$

Then there exists a unique vector $\lambda \in \mathbb{R}^k$ such that

$$\nabla f(x^*) = \lambda^\top \nabla g(x^*)$$

The function f is called *objective function*, g is called *constraint* and λ is called *Lagrange Multiplier.* ♢

References

[1] H. Fink, S. Geissel, J. Herbinger, and F. T. Seifried. “Portfolio Optimization with Optimal Expected Utility Risk Measures”. In: (2019). URL: https://ssrn.com/abstract=3412529.

[2] C. L. a. Jean-Baptiste Hiriart-Urruty. *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods.* 1st ed. Grundlehren der mathematischen Wissenschaften №306. Springer, 1993. ISBN: 9783642081620; 3642081622; 9783662064092; 366206409X. URL: libgen.li/file.php?md5=4eac7d94d7664fa3dad900a9410c48e5.

[3] D. G. Luenberger. *Linear and Nonlinear Programming.* 2nd. See Chapter 5 for the Lagrange multiplier theorem and applications. Addison-Wesley, 1984.

[4] H. Markowitz. “Portfolio Selection”. In: *The Journal of Finance* 7.1 (1952), pp. 77–91.

[5] W. F. Sharpe. “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk”. In: *The Journal of Finance* 19.3 (1964), pp. 425–442.

[9] see e.g. https://en.wikipedia.org/wiki/Lagrange_multiplier

List of Symbols

$A = \mathbf{1}^\top \Sigma^{-1}\mu \in \mathbb{R}$ auxilliary quantity in Markowitz optimization, page 9

$A_i \in \mathbb{R}$ abnormal return of asset i, page 28

$\alpha_i \in \mathbb{R}$ alpha of asset i, page 28

$\alpha_P \in \mathbb{R}$ alpha of a portfolio, page 33

$B = \mathbf{1}^\top \Sigma^{-1}\mathbf{1} \in \mathbb{R}$ auxilliary quantity in Markowitz optimization, page 9

$\beta_i \in \mathbb{R}$ beta of asset i, page 28

$\beta_P \in \mathbb{R}$ beta of a portfolio, page 33

$C = \mu^\top \Sigma^{-1}\mu \in \mathbb{R}$ auxilliary quantity in Markowitz optimization, page 9

c_t dividend yield, page 4

R_t cashflows of an asset, page 4

$d \in \mathbb{N}$ number of risky assets, page 5

$D = AC - B^2 \in \mathbb{R}$ auxilliary quantity in Markowitz optimization, page 9

$\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^d$ a vector, where all entries are 1, page 9

ε_i idiosyncratic movement of asset i, page 28

$IR_P \in \mathbb{R}$ information ratio of a portfolio, page 33

$\mu \in \mathbb{R}^d$ expected asset returns, page 5

$\mu_e = \mu - \mu_f\,\mathbf{1} \in \mathbb{R}^d$ vector of excess returns, page 20

$\mu_f \in \mathbb{R}$ expected risk-free return, page 19

$\mu_{\text{mvp}} \in \mathbb{R}$ return of minimal variance portfolio, page 12

$\bar{\mu}_P \in \mathbb{R}$ target portfolio returns (across risky and risk-free assets), page 19

$\mu_T \in \mathbb{R}$ return of the tangency portfolio, page 22

$r = (r^{(1)}, \ldots, r^{(d)})$ random vector of asset returns (in percent), page 5

$\mathbb{E}[r_i \mid r_m]$ conditionally expected returns r_i of asset i given market returns r_m, page 30

$R_P \in \mathbb{R}$ total returns of a portfolio, page 7

$r_p \in \mathbb{R}$ rate of returns of a portfolio, page 7

r_t rate of return, page 4

R_t total holding period returns of an asset (including cashflows), page 4

R'_t time series of capital gains/losses , page 4

r'_t rate of capital gain/loss, page 4

$S = (S^{(1)}, \ldots, S^{(d)})$ universe of assets, page 5

$S = \sqrt{\mu_e^\top \Sigma^{-1}\mu_e} \in \mathbb{R}$ Sharpe ratio, slope of CAL, page 20

$\Sigma \in \mathbb{R}^{d\times d}$ covariance matrix of asset returns, page 5

$\sigma^2_{\epsilon_i} \in \mathbb{R}$ non-systematic variance of asset i, page 30

$\sigma^2_i \in \mathbb{R}$ total variance of asset i, page 30

$\sigma^2_{\text{mvp}} \in \mathbb{R}$ variance of minimal variance portfolio, page 12

$\mu_P^{\blacktriangle} = \mu_P^{\blacktriangle}(\sigma_P) \in \mathbb{R}$ return of return maximal portfolio given target variance when risk-free asset included, page 21

$\sigma_P^{\blacktriangledown} = \sigma_P^{\triangledown}(\bar{\mu}_P) \in \mathbb{R}$ volatility of variance minimal portfolio given target return when risk-free asset included, page 20

$\sigma_P^{\triangledown} = \sigma_P^{\triangledown}(\mu_P) \in \mathbb{R}$ volatility of variance minimal portfolio given target return, page 9

$\sigma_T \in \mathbb{R}$ volatility of the tangency portfolio, page 22

$(S_t)_{0 \leq t \leq n}$ time series of asset prices, page 4

$TR_P \in \mathbb{R}$ Treynor ratio of a portfolio, page 33

$u \in \mathbb{R}^d$ vector of asset units of a portfolio, page 7

$w \in \mathbb{R}^d$ vector of asset weights of a portfolio, page 6

$w_f \in \mathbb{R}$ weight in risk-free asset, page 19

$w_{\blacktriangledown} \in \mathbb{R}^d$ risky asset weights on variance minimal portfolio weights given target return when risk-free asset included, page 20

$w_{\blacktriangle} \in \mathbb{R}^d$ risky asset weights on return maximal portfolio given target variance when risk-free asset included, page 21

$w_{\triangledown} \in \mathbb{R}^d$ variance minimal portfolio weights given target return, page 9

$w_{\text{mvp}} \in \mathbb{R}^d$ weight vector of minimal variance portfolio, page 12

$\mu_P^{\triangle} \in \mathbb{R}$ returns of return maximal portfolio weights given target variance, page 14

$w^{\star} \in \mathbb{R}^d$ return maximal portfolio weights given target variance, page 14

$w_T \in \mathbb{R}^d$ weights of the tangency portfolio, page 22

$X_i = r_i - \mu_f$ excess return of asset i, page 28

$X_P = r_P - \mu_f$ excess return of portfolio P, page 28