Mean-Variance Optimization and Beyond
Kaiwen Zhou

Contents

1 Introduction

This research paper is a study I've conducted on various portfolio allocation topics. It begins with an exploration of the classical mean-variance optimization framework and its derivatives, such as the CAPM theory and Arbitrage Pricing Theory. The study then delves into several models designed to address the limitations of the classical mean-variance optimization framework, including the Black-Litterman model.

The discussion continues with the Black-Litterman-APT model, which proposes a kind of 'systematic' trading strategy. This model counters the drawback of the classical Black-Litterman framework which relies heavily on personal perception and views of the market, thereby limiting the number of assets an investor can incorporate into their portfolio. After all, it's impossible for one person to have insightful knowledge about every stock in the market unless they are the market themselves. This model's origin can be traced back to paper [8].

Finally, we introduce the framework proposed by Attilio Meucci, which, in my view, is an excellent model. His insights on market invariants are enlightening, and his framework offers a structured way of making investment decisions compatible with all asset classes and various popular strategies in the field.

To bridge theory and practice, we conduct an empirical analysis comparing various models, frameworks, and techniques that we have presented and analyzed. We conclude our research with a discussion on future potential improvements to refine our current model.

This research paper should reflect my learning journey and my evolving understanding of the topics discussed. I would like to extend my heartfelt gratitude to Prof. Jonathan Goodman, whose guidance and book recommendations have significantly enriched my research experience. Indeed, there is no greater joy than having the wisdom of a mentor lighting the path of inquiry. Thanks.

2   Preliminaries

In this section, we introduce several preliminaries that would be essential for our research.

2.1   Karush−Kuhn−Tucker (KKT) conditions

The KKT conditions are the first-order necessary conditions for a non-linear constrained optimization problem.

2.1.1   Problem Formulation

Consider the following nonlinear minimization / maximization problem:

Problem 2.1. (Target Nonlinear Optimization Problem)

$$\begin{aligned} \text{optimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & g_i(\boldsymbol{x}) \leq 0, i = 1, \dots, m \\ & h_j(\boldsymbol{x}) = 0, j = 1, \dots, \ell. \end{aligned}$$

where

- $\boldsymbol{x} \in \mathbf{X}$ is the optimization variable chosen from a convex subset of $\mathbb{R}^n$.

- $f : \mathbb{R}^n \to \mathbb{R}$ is the objective function.

- $g_i : \mathbb{R}^n \to \mathbb{R}(i = 1, \dots, m)$ are the inequality constraint functions and have subderivatives at a point $x^* \in \mathbb{R}^n$..

- $h_j : \mathbb{R}^n \to \mathbb{R}(j = 1, \dots, \ell)$ are the equality constraint functions and have subderivatives at a point $x^* \in \mathbb{R}^n$.

2.1.2   Corresponding Lagrangian

Corresponding to the constrained optimization problem one can form the Lagrangian function

$$\mathcal{L}(\boldsymbol{x}, \mu, \lambda) = f(\boldsymbol{x}) + \mu^\top \mathbf{g}(\boldsymbol{x}) + \lambda^\top \mathbf{h}(\boldsymbol{x})$$

where

$$\mathbf{g}(\boldsymbol{x}) = \begin{bmatrix} g_1(\boldsymbol{x}) \\ \vdots \\ g_i(\boldsymbol{x}) \\ \vdots \\ g_m(\boldsymbol{x}) \end{bmatrix}, \quad \mathbf{h}(\boldsymbol{x}) = \begin{bmatrix} h_1(\boldsymbol{x}) \\ \vdots \\ h_j(\boldsymbol{x}) \\ \vdots \\ h_\ell(\boldsymbol{x}) \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_i \\ \vdots \\ \mu_m \end{bmatrix}, \quad \text{and} \quad \lambda = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_j \\ \vdots \\ \lambda_\ell \end{bmatrix}.$$

2.1.3   KKT Conditions

If x^* is a local optimum and the optimization problem satisfies some regularity conditions , then there exist constants $\mu_i, i = 1, \ldots, m$ and $\lambda_j, j = 1, \ldots, \ell$, called KKT multipliers, such that the following four groups of conditions hold:

- **Stationarity:**

  (a) For minimizing $f(x) : \partial f(x^*) + \sum_{j=1}^{\ell} \lambda_j \partial h_j(x^*) + \sum_{i=1}^{m} \mu_i \partial g_i(x^*) = \mathbf{0}.$

  (b) For maximizing $f(x) : -\partial f(x^*) + \sum_{j=1}^{\ell} \lambda_j \partial h_j(x^*) + \sum_{i=1}^{m} \mu_i \partial g_i(x^*) = \mathbf{0}.$

- **Primal Feasibility:**

$$h_j(x^*) = 0, \quad \text{for } j = 1, \ldots, \ell$$
$$g_i(x^*) \leq 0, \quad \text{for } i = 1, \ldots, m$$

- **Dual Feasibility:**

$$\mu_i \geq 0, \quad \text{for } i = 1, \ldots, m$$

- **Complementary Slackness:**

$$\mu_i g_i(x^*) = 0, \quad \text{for } i = 1, \ldots, m$$

Remark 2.2.

1. For **Dual Feasibility**, only the KKT multiplier attached to the nonlinear constraint functions are required to be greater than 0.

2. You can always add extra conditions "$\lambda_j h_j(x^*) = 0, \quad \text{for } j = 1, \ldots, \ell$" to the **Complementary Slackness** condition for memorization issues, because this it already true given the requirements for h_j in the **Primal Feasibility** condition.

3. In the particular case $m = 0$, i.e., when there are no inequality constraints, the KKT conditions turn into the Lagrange conditions, and the KKT multipliers are called Lagrange multipliers.

$\triangle$

2.2   Woodbury Matrix Identity

Lemma 2.3. *(Useful Lemma 1)* Suppose $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$, then we have

$$(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1}(\mathbf{I} + \mathbf{B}\mathbf{A}^{-1})^{-1} = (\mathbf{I} + \mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{A}^{-1}$$

Proof. We have

$$\mathbf{I} = (\mathbf{A} + \mathbf{B})(\mathbf{A} + \mathbf{B})^{-1}$$
$$= (\mathbf{I} + \mathbf{B}\mathbf{A}^{-1})\mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}$$
$$\implies (\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1}(\mathbf{I} + \mathbf{B}\mathbf{A}^{-1})^{-1}$$

and

$$\mathbf{I} = (\mathbf{A} + \mathbf{B})^{-1}(\mathbf{A} + \mathbf{B})$$
$$= (\mathbf{A} + \mathbf{B})^{-1}\mathbf{A}(\mathbf{I} + \mathbf{A}^{-1}\mathbf{B})$$
$$\implies (\mathbf{A} + \mathbf{B})^{-1} = (\mathbf{I} + \mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{A}^{-1}$$

Therefore, we must have

$$(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1}(\mathbf{I} + \mathbf{B}\mathbf{A}^{-1})^{-1} = (\mathbf{I} + \mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{A}^{-1}$$

$\square$

Remark 2.4. Useful in the upcoming matrix computations. $\triangle$

Lemma 2.5. *(Useful Lemma 2) Suppose* $\mathbf{I}, \mathbf{P} \in \mathbb{R}^{n \times n}$, *then we have*

$$(\mathbf{I} + \mathbf{P})^{-1} = \mathbf{I} - (\mathbf{I} + \mathbf{P})^{-1}\mathbf{P}$$

Proof. We have

$$(\mathbf{I} + \mathbf{P})^{-1} = (\mathbf{I} + \mathbf{P})^{-1}(\mathbf{I} + \mathbf{P} - \mathbf{P}) = \mathbf{I} - (\mathbf{I} + \mathbf{P})^{-1}\mathbf{P}$$

$\square$

Remark 2.6. Useful in the upcoming matrix computations. $\triangle$

Lemma 2.7. *(Useful Lemma 3) Suppose we have matrix* $\mathbf{P}, q$, *singular or non-singular, are such that* $\mathbf{I} + \mathbf{P}q$ *is non-singular, then we have*

$$(\mathbf{I} + \mathbf{P}q)^{-1}\mathbf{P} = \mathbf{P}(\mathbf{I} + q\mathbf{P})^{-1}$$

Proof. Since $\mathbf{I} + \mathbf{P}q$ is non-singular and $det(\mathbf{I} + \mathbf{P}q) = det(\mathbf{I} + q\mathbf{P}), \mathbf{I} + q\mathbf{P}$ is also non-singular. We then have

$$\mathbf{P} + \mathbf{P}q\mathbf{P} = (\mathbf{I} + \mathbf{P}q)\mathbf{P} = \mathbf{P}(\mathbf{I} + q\mathbf{P})$$
$$\implies (\mathbf{I} + \mathbf{P}q)^{-1}\mathbf{P} = \mathbf{P}(\mathbf{I} + q\mathbf{P})^{-1}$$

$\square$

Remark 2.8. Useful in the upcoming matrix computations. $\triangle$

Theorem 2.9. *(Binomial Inverse Theorem — Woodbury Matrix Identity) If* $\mathbf{A}, \mathbf{U}, \mathbf{B}, \mathbf{V}$ *are matrices of sizes* $n \times n, n \times k, k \times k, k \times n$, *respectively, then*

   (a) If $\mathbf{A}$ *and* $\mathbf{B} + \mathbf{B}\mathbf{V}\mathbf{A}^{-1}\mathbf{U}\mathbf{B}$ *are nonsingular:*

$$(\mathbf{A} + \mathbf{U}\mathbf{B}\mathbf{V})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\mathbf{B}\left(\mathbf{B} + \mathbf{B}\mathbf{V}\mathbf{A}^{-1}\mathbf{U}\mathbf{B}\right)^{-1}\mathbf{B}\mathbf{V}\mathbf{A}^{-1}$$

(b) And (a) can be simplified to

$$(\mathbf{A} + \mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\left(\mathbf{B}^{-1} + \mathbf{VA}^{-1}\mathbf{U}\right)^{-1}\mathbf{VA}^{-1}.$$

Proof.

(a)   (i) **Method 1:** Prove by verification.

First notice that

$$(\mathbf{A} + \mathbf{UBV})\mathbf{A}^{-1}\mathbf{UB} = \mathbf{UB} + \mathbf{UBVA}^{-1}\mathbf{UB} = \mathbf{U}\left(\mathbf{B} + \mathbf{BVA}^{-1}\mathbf{UB}\right).$$

Now multiply the matrix we wish to invert by its alleged inverse

$$(\mathbf{A} + \mathbf{UBV})\left(\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UB}\left(\mathbf{B} + \mathbf{BVA}^{-1}\mathbf{UB}\right)^{-1}\mathbf{BVA}^{-1}\right)$$
$$= \mathbf{I}_n + \mathbf{UBVA}^{-1} - \mathbf{U}\left(\mathbf{B} + \mathbf{BVA}^{-1}\mathbf{UB}\right)\left(\mathbf{B} + \mathbf{BVA}^{-1}\mathbf{UB}\right)^{-1}\mathbf{BVA}^{-1}$$
$$= \mathbf{I}_n + \mathbf{UBVA}^{-1} - \mathbf{UBVA}^{-1}$$
$$= \mathbf{I}_n$$

which verifies that it is the inverse.

(ii) **Method 2:** Prove by applying matrix identities.

First, in Lemma 2.5, set $\mathbf{P} = \mathbf{A}^{-1}\mathbf{X}$, we have

$$(\mathbf{I} + \mathbf{A}^{-1}\mathbf{X})^{-1} = \mathbf{I} - (\mathbf{I} + \mathbf{A}^{-1}\mathbf{X})^{-1}\mathbf{A}^{-1}\mathbf{X} \qquad (1)$$

Then, by Lemma 2.3, we have

$$(\mathbf{A} + \mathbf{X})^{-1} = (\mathbf{I} + \mathbf{A}^{-1}\mathbf{X})^{-1}\mathbf{A}^{-1}$$
$$\textit{Equation (1)} \Longrightarrow = \left(\mathbf{I} - (\mathbf{I} + \mathbf{A}^{-1}\mathbf{X})^{-1}\mathbf{A}^{-1}\mathbf{X}\right)\mathbf{A}^{-1}$$
$$= \mathbf{A}^{-1} - (\mathbf{I} + \mathbf{A}^{-1}\mathbf{X})^{-1}\mathbf{A}^{-1}\mathbf{XA}^{-1}$$
$$= \mathbf{A}^{-1} - \mathbf{A}^{-1}(\mathbf{I} + \mathbf{XA}^{-1})^{-1}\mathbf{XA}^{-1}$$
$$\text{Set } \mathbf{X} = \mathbf{UBV} \Longrightarrow = \mathbf{A}^{-1} - \mathbf{A}^{-1}(\mathbf{I} + \mathbf{UBVA}^{-1})^{-1}\mathbf{UBVA}^{-1}$$

In Lemma 2.7, set $\mathbf{P} = \mathbf{UB}, q = \mathbf{VA}^{-1}$, we have

$$\Longrightarrow = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UB}(\mathbf{I} + \mathbf{VA}^{-1}\mathbf{UB})^{-1}\mathbf{VA}^{-1}$$
$$\mathbf{B} \text{ is invertible} \Longrightarrow = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UB}\left(\mathbf{B}^{-1}(\mathbf{B} + \mathbf{BVA}^{-1}\mathbf{UB})\right)^{-1}\mathbf{VA}^{-1}$$
$$= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UB}\left(\mathbf{B} + \mathbf{BVA}^{-1}\mathbf{UB}\right)^{-1}\mathbf{BVA}^{-1}$$

Therefore, we must have

$$(\mathbf{A} + \mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UB}\left(\mathbf{B} + \mathbf{BVA}^{-1}\mathbf{UB}\right)^{-1}\mathbf{BVA}^{-1}$$

provided $\mathbf{B}$ is non-singular, and $\mathbf{I} + q\mathbf{P} = \mathbf{I} + \mathbf{VA}^{-1}\mathbf{UB}\ \mathbf{B}$ is non-singular as it's restricted by the conditions in Lemma 2.7.

(b) Since $\mathbf{B} + \mathbf{BVA}^{-1}\mathbf{UB} = \mathbf{B}(\mathbf{I} + \mathbf{VA}^{-1}\mathbf{UB})$ is nonsingular, we must have **B is invertible**. Then the two **B** terms flanking the quantity inverse in the right-hand side can be replaced with $\left(\mathbf{B}^{-1}\right)^{-1}$, which simplifies (a) to

$$(\mathbf{A} + \mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\left(\mathbf{B}^{-1} + \mathbf{VA}^{-1}\mathbf{U}\right)^{-1}\mathbf{VA}^{-1}.$$

$\square$

Remark 2.10. This matrix identity is important in proving some properties in relation to the Black-Litterman model. $\triangle$

Corollary 2.11. *(A Useful Result for Black-Litterman Model Construction) Suppose* $0 < \tau < 1, \tau \in \mathbb{R}, \mathbf{\Sigma} \in \mathbb{R}^{n \times n}, \mathbf{P} \in \mathbb{R}^{k \times n},$ *and* $\mathbf{\Omega} \in \mathbb{R}^{k \times k}.$ *Then, we have*

$$\left[(\tau\mathbf{\Sigma})^{-1} + \mathbf{P}^\top\mathbf{\Omega}^{-1}\mathbf{P}\right]^{-1} = \tau\mathbf{\Sigma} - \tau\mathbf{\Sigma}\mathbf{P}^\top(\mathbf{\Omega} + \mathbf{P}\tau\mathbf{\Sigma}\mathbf{P}^\top)^{-1}\mathbf{P}\tau\mathbf{\Sigma}$$

Proof. In Theorem 2.9, set $\mathbf{A} = (\tau\mathbf{\Sigma})^{-1}, \mathbf{U} = \mathbf{P}^\top, \mathbf{B} = \mathbf{\Omega}^{-1}, \mathbf{V} = \mathbf{P}$, then we have

$$\left[(\tau\mathbf{\Sigma})^{-1} + \mathbf{P}^\top\mathbf{\Omega}^{-1}\mathbf{P}\right]^{-1}$$

$$= \left((\tau\mathbf{\Sigma})^{-1}\right)^{-1} - \left((\tau\mathbf{\Sigma})^{-1}\right)^{-1}\mathbf{P}^\top\left(\left(\mathbf{\Omega}^{-1}\right)^{-1} + \mathbf{P}\left((\tau\mathbf{\Sigma})^{-1}\right)^{-1}\mathbf{P}^\top\right)^{-1}\mathbf{P}\left((\tau\mathbf{\Sigma})^{-1}\right)^{-1}$$

$$= \tau\mathbf{\Sigma} - \tau\mathbf{\Sigma}\mathbf{P}^\top\left(\mathbf{\Omega} + \mathbf{P}\tau\mathbf{\Sigma}\mathbf{P}^\top\right)^{-1}\mathbf{P}\tau\mathbf{\Sigma}$$

Hence, we obtain

$$\left[(\tau\mathbf{\Sigma})^{-1} + \mathbf{P}^\top\mathbf{\Omega}^{-1}\mathbf{P}\right]^{-1} = \tau\mathbf{\Sigma} - \tau\mathbf{\Sigma}\mathbf{P}^\top(\mathbf{\Omega} + \mathbf{P}\tau\mathbf{\Sigma}\mathbf{P}^\top)^{-1}\mathbf{P}\tau\mathbf{\Sigma}$$

$\square$

2.3 Generalized Least Squares

1. **Data:**

   In standard linear regression models we observe data $\{y_i, x_{ij}\}_{i=1,\ldots,n, j=2,\ldots,k}$ on n statistical units. The response values are placed in a vector $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$, and the predictor values are placed in the design matrix $\mathbf{X} = \left(\mathbf{x}_1^\top, \ldots, \mathbf{x}_n^\top\right)$, where $\mathbf{x}_i = (1, x_{i2}, \ldots, x_{ik})$ is a vector of the k predictor variables / features (including a constant) for the ith unit.

2. **Model:**

   The model forces the conditional mean of $\boldsymbol{y}$ given $\mathbf{X}$ to be a linear function of $\mathbf{X}$, i.e. $\mathbb{E}[\boldsymbol{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$, and assumes the conditional variance of the error term $\varepsilon \in \mathbb{R}^n$ given $\mathbf{X}$ is a known non-singular covariance matrix $\mathbf{\Omega} \in \mathbb{R}^{n \times n}$. This is usually written as

   $$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon, \quad \mathbb{E}[\varepsilon|\mathbf{X}] = 0, \text{Cov}[\varepsilon|\mathbf{X}] = \mathbf{\Omega}.$$

   Here $\boldsymbol{\beta} \in \mathbb{R}^k$ is a vector of unknown constants (known as "regression coefficients") that must be estimated from the data.

3. **Loss Function:**

   Suppose b is a candidate estimate for β. Then the residual vector for b will be $y - \mathbf{X}\mathbf{b}$. The generalized least squares method estimates β by minimizing the squared Mahalanobis length of this residual vector:

   $$\hat{\beta} = \operatorname*{argmin}_{b} \quad (y - \mathbf{X}\mathbf{b})^{\top}\boldsymbol{\Omega}^{-1}(y - \mathbf{X}\mathbf{b})$$
   $$= \operatorname*{argmin}_{b} \quad y^{\top}\boldsymbol{\Omega}^{-1}y + (\mathbf{X}\mathbf{b})^{\top}\boldsymbol{\Omega}^{-1}\mathbf{X}\mathbf{b} - y^{\top}\boldsymbol{\Omega}^{-1}\mathbf{X}\mathbf{b} - (\mathbf{X}\mathbf{b})^{\top}\boldsymbol{\Omega}^{-1}y,$$

   where the last two terms evaluate to scalars, resulting in

   $$\hat{\beta} = \operatorname*{argmin}_{b} \quad y^{\top}\boldsymbol{\Omega}^{-1}y + \mathbf{b}^{\top}\mathbf{X}^{\top}\boldsymbol{\Omega}^{-1}\mathbf{X}\mathbf{b} - 2\mathbf{b}^{\top}\mathbf{X}^{\top}\boldsymbol{\Omega}^{-1}y$$

4. **Solve for the GLS Estimator:**

   This objective is a quadratic form in b. Taking the gradient of this quadratic form with respect to b and equating it to zero (when $\mathbf{b} = \hat{\beta}$) gives

   $$2\mathbf{X}^{\top}\boldsymbol{\Omega}^{-1}\mathbf{X}\hat{\beta} - 2\mathbf{X}^{\top}\boldsymbol{\Omega}^{-1}y = 0$$

   Therefore, the minimum of the objective function can be computed yielding the explicit formula:
   $$\hat{\beta}_{GLS} = \left(\mathbf{X}^{\top}\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\boldsymbol{\Omega}^{-1}y$$

   The quantity $\boldsymbol{\Omega}^{-1}$ is known as the precision matrix (or dispersion matrix), a generalization of the diagonal weight matrix.

2.3.1 Properties of GLS Estimator

The GLS estimator is

- unbiased,

- consistent,

- efficient,

- asymptotically normal

- $\mathbb{E}[\hat{\beta}|\mathbf{X}] = \beta$ and $\operatorname{Cov}[\hat{\beta}|\mathbf{X}] = \left(\mathbf{X}^{\top}\Omega^{-1}\mathbf{X}\right)^{-1}$.

Property 2.12. *(Understanding GLS) GLS is equivalent to applying ordinary least squares to a linearly transformed version of the data.*

Solution: To see this, factor $\Omega = \mathbf{C}\mathbf{C}^\top$ using the Cholesky decomposition. Then if we pre-multiply both sides of the equation $y = \mathbf{X}\beta + \varepsilon$ by $\mathbf{C}^{-1}$, i.e.

$$y = \mathbf{X}\beta + \varepsilon \quad \Longrightarrow \quad \mathbf{C}^{-1}y = \mathbf{C}^{-1}\mathbf{X}\beta + \mathbf{C}^{-1}\varepsilon$$

we get an equivalent linear model

$$y^* = \mathbf{X}^*\beta + \varepsilon^*$$

where $y^* = \mathbf{C}^{-1}y$, $\mathbf{X}^* = \mathbf{C}^{-1}\mathbf{X}$, and $\varepsilon^* = \mathbf{C}^{-1}\varepsilon$.

In this model

$$
\begin{aligned}
Cov = \mathbb{V}\left[\varepsilon^* | \mathbf{X}\right] &= \mathbb{E}\left[(\varepsilon^* - \mathbb{E}[\varepsilon^*])(\varepsilon^* - \mathbb{E}[\varepsilon^*])^\top | \mathbf{X}\right] \\
&= \mathbb{E}\left[C^{-1}(\varepsilon - \mathbb{E}[\varepsilon])(\varepsilon - \mathbb{E}[\varepsilon])^\top (C^{-1})^\top | \mathbf{X}\right] \\
&= \mathbf{C}^{-1}\mathbb{V}[\varepsilon | \mathbf{X}]\left(\mathbf{C}^{-1}\right)^\top \\
&= \mathbf{C}^{-1}\Omega\left(\mathbf{C}^{-1}\right)^\top \\
&= \mathbf{I}
\end{aligned}
$$

where $\mathbf{I}$ is the identity matrix.

Thus we can efficiently estimate β by applying Ordinary least squares (OLS) to the transformed data, which requires minimizing

$$\left(y^* - \mathbf{X}^*\beta\right)^\top \left(y^* - \mathbf{X}^*\beta\right) = (y - \mathbf{X}\mathbf{b})^\top \Omega^{-1}(y - \mathbf{X}\mathbf{b})$$

This has the effect of standardizing the scale of the errors and "de-correlating" them. Since OLS is applied to data with homoscedastic errors, the GaussMarkov theorem applies, and therefore the GLS estimate is the best linear unbiased estimator for β.

2.4   Bayesian Statistical Model

A Bayesian statistical model consists of:

1. An unknown target parameter $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^\ell$.

2. A vector-valued random variable $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^d$ where realizations of $\boldsymbol{x}$ have been observed and known.

3. $\boldsymbol{x}$ is assumed to be distributed according to the likelihood function $f(\boldsymbol{x}|\boldsymbol{\theta})$. After conditioning on $\boldsymbol{\theta}$, $f(\boldsymbol{x}|\boldsymbol{\theta})$ forms a density on the data space $\mathcal{X} \subseteq \mathbb{R}^d$.

4. A prior density $\pi(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$ on Θ.

5. The posterior is the density $p(\boldsymbol{\theta}|\boldsymbol{x})$ on Θ proportional to $f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$, i.e.

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = Cf(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$$

and the normalization factor C can be calculated using the fact that $\int p(\boldsymbol{\theta}|\boldsymbol{x}) = 1$.

Intuition: The Bayesian framework can be viewed as a journey of finding the right distribution for $\boldsymbol{\theta}$. Initially, you have a rough idea of what the distribution of $\boldsymbol{\theta}$ might be, which is called the prior, $\pi(\boldsymbol{\theta})$. Next, you use the "real" observations of $\boldsymbol{x}$ to update your understanding of the latent parameter $\boldsymbol{\theta}$'s distribution, denoted by $p(\boldsymbol{\theta}|\boldsymbol{x})$. In the case of $\ell = d = 1$, we have

$$p(\theta|X_1 = x_1, \ldots, X_n = x_n) \propto f(X_1 = x_1, \ldots, X_n = x_n|\theta)\,\pi(\theta)$$

As more and more observations are provided, you will eventually find the true distribution of $\boldsymbol{\theta}$, or a distribution around the true value of $\boldsymbol{\theta}$ with a narrow interval.

Remark 2.13.

1. In Bayesian statistics, all statistical inference is based on the posterior $p(\boldsymbol{\theta}|\boldsymbol{x})$.

2. In the Bayesian context, p is used more specifically to denote these probability distributions related to Bayesian inference, whereas the letter f is more general and can represent any probability density or mass function, whether in a Bayesian or frequentist context.

3. When using the likelihood function, we can drop its coefficient, which essentially means that only the relative value of $f(\boldsymbol{x}|\boldsymbol{\theta})$ matters.

   For instance, if $x|\theta \sim \mathcal{N}(0,1)$, then $f(x|\theta) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. We could use $f'(x|\theta) = e^{-x^2/2}$ instead, because the scale, i.e., $c = \frac{1}{\sqrt{2\pi}}$, is not important.

   This occurs because when maximizing the likelihood function to solve for the MLE estimator, both $c \cdot f(\boldsymbol{x}|\boldsymbol{\theta})$ and $f'(\boldsymbol{x}|\boldsymbol{\theta})$ will have the same maximum. Additionally, when calculating the normalization factor C in Bayesian statistics, both $c \cdot f(\boldsymbol{x}|\boldsymbol{\theta})$ and $f'(\boldsymbol{x}|\boldsymbol{\theta})$ will produce the same posterior due to the way we compute the posterior. As a result, we do not concern ourselves with the scale.

   $\triangle$

Theorem 2.14. *(A Useful Theorem) If a multivariate normal random variable $\boldsymbol{\theta} \in \mathbb{R}^k$ has density $p(\boldsymbol{\theta})$ and*

$$-2\log p(\boldsymbol{\theta}) = \boldsymbol{\theta}^\top \mathbf{M}\boldsymbol{\theta} - 2\boldsymbol{u}^\top \boldsymbol{\theta} + (\textit{ terms without } \boldsymbol{\theta})$$

where $\mathbf{M} \in \mathbb{R}^{k \times k}$ is symmetric and invertible, then

$$\mathbb{V}[\boldsymbol{\theta}] = \mathbf{M}^{-1} \textit{ and } \mathbb{E}[\boldsymbol{\theta}] = \mathbf{M}^{-1}\boldsymbol{u}.$$

Solution: For M symmetric and invertible, we have

$$
\begin{aligned}
-2\log p(\boldsymbol{\theta}) &= \boldsymbol{\theta}^\top \mathbf{M}\boldsymbol{\theta} - 2\boldsymbol{u}^\top\boldsymbol{\theta} + (\text{ terms without } \boldsymbol{\theta}) \\
&= \boldsymbol{\theta}^\top \mathbf{M}\boldsymbol{\theta} - 2\boldsymbol{u}^\top \mathbf{M}^{-1}\mathbf{M}\boldsymbol{\theta} + (\text{ terms without } \boldsymbol{\theta}) \\
&= (\boldsymbol{\theta} - \mathbf{M}^{-1}\boldsymbol{u})^\top \mathbf{M}(\boldsymbol{\theta} - \mathbf{M}^{-1}\boldsymbol{u}) - \boldsymbol{u}^\top \mathbf{M}^{-1}\boldsymbol{u} + (\text{ terms without } \boldsymbol{\theta}) \\
&= (\boldsymbol{\theta} - \mathbf{M}^{-1}\boldsymbol{u})^\top \mathbf{M}(\boldsymbol{\theta} - \mathbf{M}^{-1}\boldsymbol{u}) + (\text{ terms without } \boldsymbol{\theta})
\end{aligned}
$$

Since $p(\boldsymbol{\theta})$ is a density function, we msut have

$$
\begin{aligned}
1 = \int p(\boldsymbol{\theta})d\boldsymbol{\theta} &= \int \exp\left[-\frac{1}{2}\left((\boldsymbol{\theta} - \mathbf{M}^{-1}\boldsymbol{u})^\top \mathbf{M}(\boldsymbol{\theta} - \mathbf{M}^{-1}\boldsymbol{u}) + (\text{ terms without } \boldsymbol{\theta})\right)\right]d\boldsymbol{\theta} \\
&= \sqrt{(2\pi)^k|\mathbf{M}^{-1}|} \cdot \exp\left[-\frac{1}{2}\left((\text{ terms without } \boldsymbol{\theta})\right)\right] \\
\implies \frac{1}{\sqrt{(2\pi)^k|\mathbf{M}^{-1}|}} &= \exp\left[-\frac{1}{2}\left((\text{ terms without } \boldsymbol{\theta})\right)\right]
\end{aligned}
$$

Hence, we get

$$
\begin{aligned}
p(\boldsymbol{\theta}) &= \exp\left[-\frac{1}{2}\left((\boldsymbol{\theta} - \mathbf{M}^{-1}\boldsymbol{u})^\top \mathbf{M}(\boldsymbol{\theta} - \mathbf{M}^{-1}\boldsymbol{u}) + (\text{ terms without } \boldsymbol{\theta})\right)\right] \\
&= \frac{1}{\sqrt{(2\pi)^k|\mathbf{M}^{-1}|}}\exp\left[-\frac{1}{2}\left((\boldsymbol{\theta} - \mathbf{M}^{-1}\boldsymbol{u})^\top \mathbf{M}(\boldsymbol{\theta} - \mathbf{M}^{-1}\boldsymbol{u})\right)\right]
\end{aligned}
$$

This is the density function for a multi-variable Gaussian distribution $\mathcal{N}\left(\mathbf{M}^{-1}\boldsymbol{u}, \mathbf{M}^{-1}\right)$; therefore, we can view parameter $\boldsymbol{\theta}$ as

$$
\boldsymbol{\theta} \sim \mathcal{N}\left(\mathbf{M}^{-1}\boldsymbol{u}, \mathbf{M}^{-1}\right)
$$

Then, we must have

$$
\mathbb{V}[\boldsymbol{\theta}] = \mathbf{M}^{-1} \text{ and } \mathbb{E}[\boldsymbol{\theta}] = \mathbf{M}^{-1}\boldsymbol{u}.
$$

2.5 Multivariate log-normal

If $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a multivariate normal distribution, then $Y_i = \exp(X_i)$ has a multivariate log-normal distribution. The exponential is applied elementwise to the random vector $\boldsymbol{X}$. The mean of $\boldsymbol{Y}$ is

$$
\mathrm{E}[\boldsymbol{Y}]_i = e^{\mu_i + \frac{1}{2}\Sigma_{ii}}
$$

and its covariance matrix is

$$
\mathrm{Var}[\boldsymbol{Y}]_{ij} = e^{\mu_i + \mu_j + \frac{1}{2}(\Sigma_{ii} + \Sigma_{jj})}\left(e^{\Sigma_{ij}} - 1\right)
$$

3   Fictional Market Sample

Motivation: We introduce the example used in the paper (He and Litterman (2002)) regarding the international equity market so as to better illustrate our point.

Data has been provided for the index volatility, equilibrium portfolio weight, and implied returns. The weighting is based on the market capitalization of each market and the expected returns are implied by the market prices.

Table 1: Market Data

Country	Equity Index Volatility (%)	Equilibrium Portfolio Weight (%)	Equilibrium Expected Returns (%)
Australia	16.0	-1.097079	3.9
Canada	20.3	18.74685	6.9
France	24.8	4.287915	8.4
Germany	27.1	2.364488	9.0
Japan	21.0	11.429539	4.3
UK	20.0	13.61645	6.8
USA	18.7	52.082372	7.6

Additionally, we provide the covariance matrix for this dataset. Lastly, we set the risk-

Table 2: Covariance Matrix of Returns

	AUS	CAN	FRA	GER	JAP	UK	USA
AUS	0.0256	0.01585	0.018967	0.02233	0.01475	0.016384	0.014691
CAN	0.01585	0.041209	0.033428	0.036034	0.027923	0.024685	0.024751
FRA	0.018967	0.033428	0.061504	0.057866	0.018488	0.038837	0.030979
GER	0.02233	0.036034	0.057866	0.073441	0.020146	0.042113	0.033092
JAP	0.01475	0.013215	0.018488	0.020146	0.0441	0.01701	0.012017
UK	0.016384	0.024685	0.038837	0.042113	0.01701	0.04	0.024385
USA	0.014691	0.029572	0.030979	0.033092	0.012017	0.024385	0.034969

free rate to be 2%; that is,

$$r_f = 2\% \tag{2}$$

4 Mean-Variance Optimization

The MVO is a single period model. We will use the fictional data in Section 3 to better illustrate our points along the way.

4.1 Financial Concepts

Definition 4.1. *(Rate of Return) We define the **rate of return (RoR or r)** as*

$$r = \text{rate of return} = \frac{\text{amount received} - \text{amount invested}}{\text{amount invested}}$$

Definition 4.2. *(Total Return) We define the **total return (total R or R)** as*

$$R = \text{total return} = \frac{\text{amount received}}{\text{amount invested}}$$

Remark 4.3.

(a) In some context, the "total return" defined above is called "gross return" and "rate of return" is called "total return".

Such definition is **consistent** for both vanilla trade and short sale. To be specific, we have

1. Suppose we are in a vanilla trade, where you invest X_0 dollars and get X_1 dollars from this investment ($X_0 > 0$). Then your RoR is $r = \frac{X_1 - X_0}{X_0}$, and total return is $R = \frac{X_1}{X_0}$. That is,

$$\text{vanilla trade} : \begin{cases} \text{pay} & X_0 \\ \text{get} & X_1 \end{cases} \implies \begin{cases} \text{RoR} = r = \frac{X_1 - X_0}{X_0} \\ R = \frac{X_1}{X_0} \end{cases}$$

2. Suppose we are shorting, where you invest $-X_0$ dollars and get $-X_1$ dollars from this investment ($X_0 > 0$). Then your RoR is $r = \frac{-X_1 - (-X_0)}{-X_0}$, and total return is $R = \frac{-X_1}{-X_0}$. That is,

$$\text{short sale} : \begin{cases} \text{pay} & -X_0 \\ \text{get} & -X_1 \end{cases} \implies \begin{cases} \text{RoR} = r = \frac{X_1 - X_0}{X_0} \\ R = \frac{X_1}{X_0} \end{cases}$$

$\triangle$

Definition 4.4. *(Relative Weighted Portfolio) Suppose we invested in n asset where the amount invested in the i-th asset is X_{0i}, $i = 1, 2, \ldots, n$, and the total capital invested is X_0. Then, we have*

$$X_0 = \sum_{i=1}^{n} X_{0i}.$$

Then, the equation above describes an allocation/portfolio.

It's usually better to rewrite the above equation as

$$X_0 = \sum_{i=1}^n w_i X_0.$$

*where $w_i = \frac{X_{0i}}{X_0}$, and $\sum_{i=1}^n w_i = 1$. We call such formulation **relative weighted portfolio**. If we allow shorting, X_{0i} can be negative, and so does the corresponding w_i.*

Remark 4.5. The reason for us to formulate the portfolio as the one with relative weights are

1. More Intuitive: It is easier to interpret a statement such as "thirty percent of one's budget is invested in xyz" than "his investment consists, among others, of a thousand shares of xyz".

2. Reduce computational complexity: Results in simpler forms for the computation in the following different allocation frameworks.

$\triangle$

Definition 4.6. *(The Total Return of The Portfolio) The total return of the portfolio is*

$$R_p := \frac{\text{amount received}}{\text{amount invested}} = \frac{\sum_{i=1}^n R_i w_i X_0}{X_0} = \sum_{i=1}^n R_i w_i.$$

Suppose we want to invest in n different assets $A_1, A_2, \ldots, A_n$ and the rate of return for each of them in one period can be viewed as a random variable $r_i, i = 1, 2, \ldots, n$. Put this in the vector format, we can rewrite it as $\boldsymbol{r} = (r_1, r_2, \ldots, r_n) \in \mathbb{R}^n$.

Definition 4.7. *(The Portfolio Rate of Return) Then, the portfolio rate of return is*

$$r_p = \frac{\text{amount received} - \text{amount invested}}{\text{amount invested}} = \frac{X_0 \left(\sum_{i=1}^n R_i w_i - \sum_{i=1}^n w_i \right)}{X_0} = \sum_{i=1}^n r_i w_i = \boldsymbol{w}^\top \boldsymbol{r}$$

where $w := (w_1, \ldots, w_n)$.

Definition 4.8. *(Expected Return of the Assets) Since each r_i can be viewed as a random variable, we can then define the expected return of assets as*

$$\boldsymbol{\mu} := \mathbb{E}[\boldsymbol{r}] \in \mathbb{R}^n$$

where $\boldsymbol{r} = (r_1, r_2, \ldots, r_n) \in \mathbb{R}^n$.

Definition 4.9. *(Expected Return of the Portfolio) Then, the expected return of the portfolio is*

$$\mu_p := \mathbb{E}\left[r_p\right] = \mathbb{E}\left[\boldsymbol{w}^\top \boldsymbol{r}\right] = \boldsymbol{w}^\top \mathbb{E}[\boldsymbol{r}] = \boldsymbol{w}^\top \boldsymbol{\mu}.$$

where $\boldsymbol{\mu} := \mathbb{E}[\boldsymbol{r}]$.

Remark 4.10.

1. Sometimes, we use the notation $\boldsymbol{\mu} := \mathbb{E}\left[\boldsymbol{r}\right]$.

2. Since all the r_i's are random variables, μ_p is a random variable.

$\triangle$

Definition 4.11. *(The Variance of the Portfolio Rate of Return) The variance of the portfolio return is given by*

$$
\begin{aligned}
\sigma_p^2 = \mathbf{V}\left[r_p\right] = \mathbb{E}\left[(r_p - \mathbb{E}\left[r_p\right])^2\right] \\
= \mathbb{E}\left[\left(\boldsymbol{w}^\top \boldsymbol{r} - \boldsymbol{w}^\top \mathbb{E}[\boldsymbol{r}]\right)^2\right] \\
= \mathbb{E}\left[\boldsymbol{w}^\top (\boldsymbol{r} - \mathbb{E}[\boldsymbol{r}])(\boldsymbol{r} - \mathbb{E}[\boldsymbol{r}])^\top \boldsymbol{w}\right] \\
= \boldsymbol{w}^\top \mathbb{E}\left[(\boldsymbol{r} - \mathbb{E}[\boldsymbol{r}])(\boldsymbol{r} - \mathbb{E}[\boldsymbol{r}])^\top\right] \boldsymbol{w} \\
= \boldsymbol{w}^\top \operatorname{var}[\boldsymbol{r}] \boldsymbol{w} \\
= \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w},
\end{aligned}
$$

where $\boldsymbol{\Sigma} := \operatorname{var}[\boldsymbol{r}]$ is the covariance matrix of returns. $\boldsymbol{\Sigma}$ is symmetric and semi-positive definite.

Definition 4.12. *(Risk Preference)*

1. *Risk Averse: Risk is bad and should be minimized.*

2. *Risk-preferring: Risk is good, like taking risks.*

3. *Risk Neutral: Don't like or hate risk.*

Definition 4.13. *(Non-satiation) An investor is Non-satiation means if everything else holds fixed, the investor always wants more money/higher return.*

Definition 4.14. *(zero-cost portfolio of risky securities) A zero-cost portfolio of risky securities is a portfolio $\boldsymbol{w}$ such that*

$$\boldsymbol{w}^\top \boldsymbol{e} = 0$$

where $\boldsymbol{e} = (1, 1, \ldots, 1) \in \mathbb{R}^n$.

Definition 4.15. *(market neutral portfolio of risky securities) A market neutral portfolio of risky securities is a portfolio $\boldsymbol{w}$ such that*

$$\boldsymbol{w}^\top \boldsymbol{\beta} = 0$$

where $\boldsymbol{\beta} := (\beta_1, \ldots, \beta_m)$ is the vector of security betas with respect to the market.

Definition 4.16. *(Market Portfolio) The market portfolio is a portfolio such that*

(a) *It consists of all securities available to investors, and*

(b) *each security in the market portfolio is held in proportion to its market value (market capitalization) relative to the total market value of all securities.*

4.2 MVO With Only Risky Securities

4.2.1 Formulation

Motivation: We want to analyze the distribution of the portfolio mean rate of return.

In the classical Mean-Variance Analysis, we need to make two basic assumptions:

Assumption 4.17. *(Core Assumption)*

1. *The investor exhibits risk aversion (Definition 4.12 and non-satiation (Definition 4.13) preferences.*

2. *The investor's initial capital is not null.*

Using the concepts and notation in Section 4.1, suppose we want to invest in n different assets $A_1, A_2, \ldots, A_n$ and the rate of return for each of them in one period can be viewed as a random variable $r_i, i = 1, 2, \ldots, n$. Put this in the vector format, we can rewrite it as $\boldsymbol{r} = (r_1, r_2, \ldots, r_n) \in \mathbb{R}^n$. From Definition 4.7, we can write the portfolio rate of return as

$$r_p = \boldsymbol{w}^\top \boldsymbol{r}$$

where $\boldsymbol{w} := (w_1, \ldots, w_n) \in \mathbb{R}^n$ is the vector of portfolio weights. Then, by Definition 4.8 and Definition 4.9, we obtain the expected return as

$$\mu_p = \boldsymbol{w}^\top \boldsymbol{\mu},$$

where $\boldsymbol{\mu} := \mathbb{E}[\boldsymbol{r}] \in \mathbb{R}^n$ is the expected return of assets. Then, by Definition 4.11, we get the variance of the portfolio return is given by

$$\sigma_p^2 = \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$$

Then, we can start to construct our MVO framework.

MVO Framework: The MVO suggests that investors should allocate their wealth by solving the one-period mean-variance optimization (MVO) problem

$$\min_{\boldsymbol{w}} \quad \frac{1}{2} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$$
$$\text{s.t.} \quad \boldsymbol{w}^\top \boldsymbol{\mu} = \mu_p, \quad \boldsymbol{w}^\top \boldsymbol{e} = 1$$

where

- $\boldsymbol{e} := (1, \ldots, 1) \in \mathbb{R}^n$,

- $\boldsymbol{w} \in \mathbb{R}^n$ is the vector of portfolio weights,

- $\boldsymbol{\mu} \in \mathbb{R}^n$ is the vector of expected returns,

- $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ is the covariance matrix of returns, and

- $\mu_p \geq 0$ is the desired level of portfolio return.

Remark 4.18.

1. Term $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ shall be treated as known and should be provided by the investor.

2. This is often called the Variance-Minimization formulation (or Risk-Minimization formulation). In such formulation, we treat variable μ_p as a fixed value. And for every fixed μ_p, there is a $\boldsymbol{w}_p$ that minimizes $\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$ which gives the minimized variance of portfolio return $\sigma_p = \sqrt{\boldsymbol{w}_p^\top \boldsymbol{\Sigma} \boldsymbol{w}_p}$.

3. We assume that $\mu_p > 0$ is chosen such that a solution to the MVO exists. In general the first constraint is frequently changed to $\boldsymbol{w}^\top \boldsymbol{\mu} \leq \mu_p$.

4. There are in several different but mathematically equivalent formulations of the MVO problem. For example:

   (The Expected Return Maximization Formulation) The expected return maximization formulation is stated as

   $$\max_{\boldsymbol{w}} \quad \boldsymbol{w}^\top \boldsymbol{\mu}$$
   $$\text{s.t.} \quad \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} = \sigma_p^2, \quad \boldsymbol{w}^\top \boldsymbol{e} = 1$$

   where σ_p^2 is the desired level of portfolio variance (risk) viewed as fixed variable.

   (The Risk Aversion Formulation) The risk aversion formulation is stated as

   $$\max_{\boldsymbol{w}} \quad \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\lambda}{2} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$$
   $$\text{s.t.} \quad \boldsymbol{w}^\top \boldsymbol{e} = 1, \lambda = \lambda_0$$

   where $\lambda \geq 0$ can be interpreted as the the risk aversion coefficient specified by the investor.

   However, this is not exactly the risk-aversion coefficient specified in the utility function. For the sake of consistency, if the risk-aversion coefficient specified in the CARA utility function is γ, then $\lambda = \gamma W_0$ where W_0 is your initial wealth. However, since γ is still a free-floating variable with no restrictions, many people abuse a little bit of this notation and claim that λ is the risk-aversion coefficient, but we really should know the subtle different underlying these two.

   $\triangle$

Now, from **??**, we obtain that with the formulation specified above, the MVO problem is a quadratic programming (QP) problem with linear equality constraints, i.*e.* $A \equiv 0$. Therefore, we can solve it in closed form. Let's now solve the MVO problem

Problem 4.19. (**Risk-Minimization formulation MVO)**

$$\min_{\boldsymbol{w}} \quad \frac{1}{2} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$$
$$\text{s.t.} \quad \boldsymbol{w}^\top \boldsymbol{\mu} = \mu_p, \quad \boldsymbol{w}^\top \boldsymbol{e} = 1$$

Solution: The Lagrangian takes the form

$$L \equiv L(\boldsymbol{w}, \lambda, \gamma) := \frac{1}{2} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} + \lambda \left(1 - \boldsymbol{w}^\top \boldsymbol{e}\right) + \gamma \left(\mu_p - \boldsymbol{w}^\top \boldsymbol{\mu}\right)$$

By the KKT condition (Section 2.1), we have

- **Stationarity:**

$$FOC = \frac{\partial L}{\partial \boldsymbol{w}} = \boldsymbol{\Sigma} \boldsymbol{w} - \lambda \boldsymbol{e} - \gamma \boldsymbol{\mu} = \boldsymbol{0}. \tag{3}$$

- **Primal Feasibility:**

$$\boldsymbol{w}^\top \boldsymbol{\mu} = \mu_p, \quad \boldsymbol{w}^\top \boldsymbol{e} = 1 \tag{4}$$

Remark 4.20. Since there is no inequality constraints for such formulation, we omitted the **Dual Feasibility** condition and the **Complementary Slackness** condition. However, there will be these two for the more general formulation as noted in Item 3. △

Form Equation (3), we get

$$\boldsymbol{w} = \boldsymbol{\Sigma}^{-1}(\lambda \boldsymbol{e} + \gamma \boldsymbol{\mu}) \tag{5}$$

Plug the result for $\boldsymbol{w}$ in Equation (5) into the two equations in Equation (4), we get

$$\boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1}(\lambda \boldsymbol{e} + \gamma \boldsymbol{\mu}) = 1, \qquad \boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e} \lambda + \boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \gamma = 1,$$
$$\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}(\lambda \boldsymbol{e} + \gamma \boldsymbol{\mu}) = \mu_p \quad \Longleftrightarrow \quad \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e} \lambda + \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \gamma = \mu_p$$

Now, for simplicity, we set $A = \boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e}$, $B = \boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e}$, and $C = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$.

Remark 4.21. We have $\boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e}$ because $\boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$ and $\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e}$ are actually real numbers and $\left(\boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}\right)^\top = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e}$. △

Then we have

$$\begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} \lambda \\ \gamma \end{bmatrix} = \begin{bmatrix} 1 \\ \mu_p \end{bmatrix} \tag{6}$$

$$\Longrightarrow \begin{bmatrix} \lambda^* \\ \gamma^* \end{bmatrix} = \begin{bmatrix} A & B \\ B & C \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \mu_p \end{bmatrix} = \frac{1}{AC - B^2} \begin{bmatrix} C & -B \\ -B & A \end{bmatrix} \begin{bmatrix} 1 \\ \mu_p \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} C - B\mu_p \\ -B + A\mu_p \end{bmatrix} \tag{7}$$

where $\Delta = AC - B^2$.

Then, plug in Equation (5), we get

$$\boldsymbol{w}^* = \lambda^* \boldsymbol{\Sigma}^{-1} \boldsymbol{e} + \gamma^* \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \tag{8}$$

This is clearly the minimum as $\frac{\partial^2 L}{\partial \boldsymbol{w}^2} = \boldsymbol{\Sigma}$ is positive definite.

We call $\boldsymbol{w}^*$ a mean-variance efficient portfolio/allocation with respect to μ_p.

4.2.2   Understanding the Optimal Allocation

Motivation: We want to find the relationship between μ_p and $\sigma_p = \sqrt{{\boldsymbol{w}^*}^\top \boldsymbol{\Sigma} \boldsymbol{w}^*}$ for the corresponding mean-variance efficient portfolio $\boldsymbol{w}^*$.

By construction, the expected return of the optimal portfolio is

$$\mu_p = {\boldsymbol{w}^*}^\top \boldsymbol{\mu}$$

Some algebra shows that the variance of the optimal portfolio is

$$
\begin{aligned}
\sigma_p^{*2} &= {\boldsymbol{w}^*}^\top \boldsymbol{\Sigma} \boldsymbol{w}^* \\
&= \left(\lambda^* \boldsymbol{\Sigma}^{-1} \boldsymbol{e} + \gamma^* \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}\right)^\top \boldsymbol{\Sigma} \left(\lambda^* \boldsymbol{\Sigma}^{-1} \boldsymbol{e} + \gamma^* \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}\right) \\
&= \lambda^{*2} \boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e} + 2\lambda^* \gamma^* \boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \gamma^{*2} \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\
&= \lambda^* (\lambda^* \boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e} + \gamma^* \boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}) + \gamma^* (\lambda^* \boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \gamma^* \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}) \\
&= \lambda^* (\lambda^* A + \gamma^* B) + \gamma^* (\lambda^* B + \gamma^* C) \\
\textit{Equation (6)} \implies &= \lambda^* \cdot 1 + \gamma^* \cdot \mu_p \\
\textit{Equation (7)} \implies &= \frac{1}{\Delta}(C - B\mu_p) + \frac{1}{\Delta}(-B + A\mu_p)\mu_p \\
&= \frac{A\mu_p^2 - 2B\mu_p + C}{\Delta},
\end{aligned}
$$

Hence, we have

$$\sigma_p^{*2} = \frac{A\mu_p^2 - 2B\mu_p + C}{\Delta} \tag{9}$$

which is a quadratic function of μ_p.

Remark 4.22.

1. If we let μ_p vary, the efficient portfolios $\boldsymbol{w}^*$ form a a **hyperbola** in the $(\sigma^* (\mu_p), \mu_p)$-plane and a **parabola** in the $\left(\sigma^{*2} (\mu_p), \mu_p\right)$-plane.

2. This hyperbola is called **the efficient frontier.**

$\triangle$

We treat the Equilibrium Expected Returns (%) in Table 1 as the expected returns $\boldsymbol{\mu}$ and the covariance matrix (Table 2) as $\boldsymbol{\Sigma}$, we get

4.2.3   The Global Minimum Variance Portfolio

The global minimum variance portfolio (GMV) is the portfolio that has the lowest variance out of all efficient portfolio. It's pretty crucial to our interpretation of the MVO framework and also the two fund separation theorem.
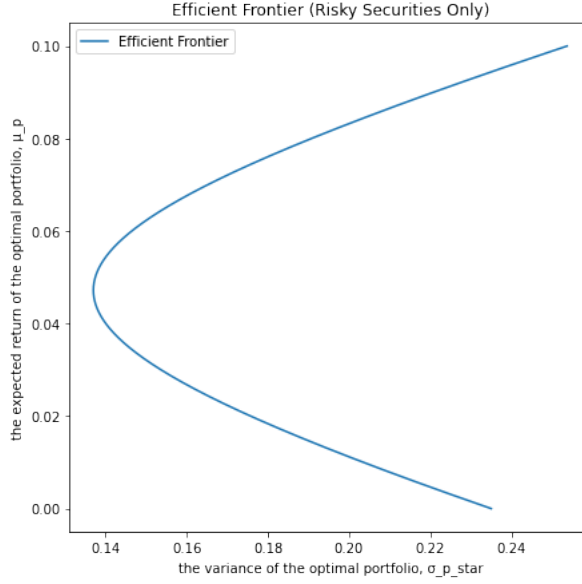
Figure 1: Efficient Frontier (Risky Securities Only)

We find it by solving

$$\frac{d\sigma_p^{*^2}}{d\mu_p} = \frac{2A\mu_p - 2B}{\Delta} \equiv 0$$

Therefore $A\mu_p = B$ so that $\gamma^* = \frac{\mu_p A - B}{\Delta} = 0$ and

$$\boldsymbol{w}_{gmv}^* = \lambda^* \boldsymbol{\Sigma}^{-1} \boldsymbol{e} = \frac{C - \mu_p B}{AC - B^2} \boldsymbol{\Sigma}^{-1} \boldsymbol{e} = \frac{\boldsymbol{\Sigma}^{-1} \boldsymbol{e}}{\boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e}}. \tag{10}$$

Remark 4.23. We observe that $\boldsymbol{w}_{gmv}^*$ does not depend on the expected returns $(\boldsymbol{\mu})$ of the risky securities. $\triangle$

4.2.4 Two Fund Separation: Only Risky Securities

Motivation: We want to see if any portfolio on the efficient frontier can be represented by a linear combination of two special portfolios on the efficient frontier. That way, we can attain any efficient portfolio by investing in a combination of two special portfolios (funds). Intuitively, one of the special portfolios could be the global minimum variance portfolio $\boldsymbol{w}_{gmv}^*$. Now, we want to find the other one by a rather witty construction.

Recall that for all minimum-variance portfolios where $\lambda^* = \lambda^* (\mu_p)$ and $\gamma^* = \gamma^* (\mu_p)$, we have

$$w^* = \lambda^* \boldsymbol{\Sigma}^{-1} \boldsymbol{e} + \gamma^* \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$$
$$= \lambda^* A \boldsymbol{w}_{gmv}^* + \gamma^* \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$$
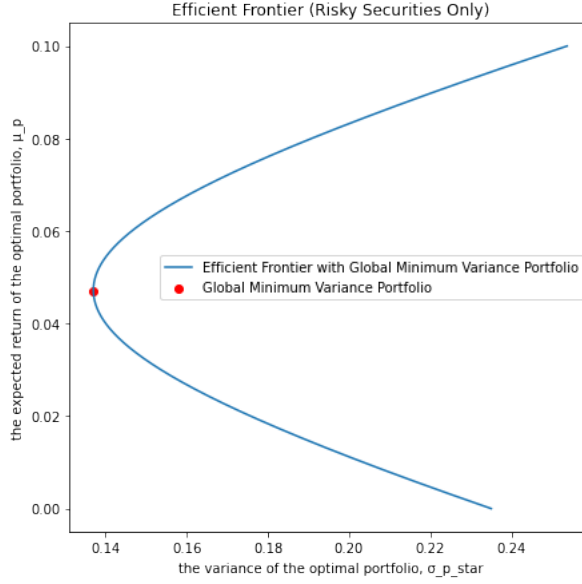
Figure 2: Efficient Frontier (Risky Securities Only) with Global Minimum Variance Portfolio

We want to rewrite $\gamma^* \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} = \gamma^* T \boldsymbol{w}_d^*, T \in \mathbb{R}$. Then, we must have

$$\boldsymbol{e}^\top \boldsymbol{w}_d^* = 1, \quad T\boldsymbol{w}_d^* = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$$

Inspired by the look of $\boldsymbol{w}_{gmv}^*$, we set

$$\boldsymbol{w}_d^* = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}$$

Then we must have

$$\boldsymbol{w}^* = \lambda^* A \boldsymbol{w}_{gmv}^* + \gamma^* B \boldsymbol{w}_d^*$$

where $A = \boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e}$ and $B = \boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e}$.

From Equation (6), it's easy to show that $\lambda^* A + \gamma^* B = 1$, which implies that any portfolio on the efficient frontier can be represented by a linear combination of $\boldsymbol{w}_{gmv}^*$ and $\boldsymbol{w}_d^*$.

Remark 4.24. Some might ask why the efficient frontier is not a line crossing the red and orange points. This is because the Two Funds Separation Theorem states that it is the weights that can be represented by a linear combination of the two funds, not the return and volatility. △

4.3 MVO with Risk-free Securities

4.3.1 Formulation

A risk-free asset has a return that is deterministic (that is; known as certainty). We will denote the return of the risk-less security by r_f. We observe that for the risk-less
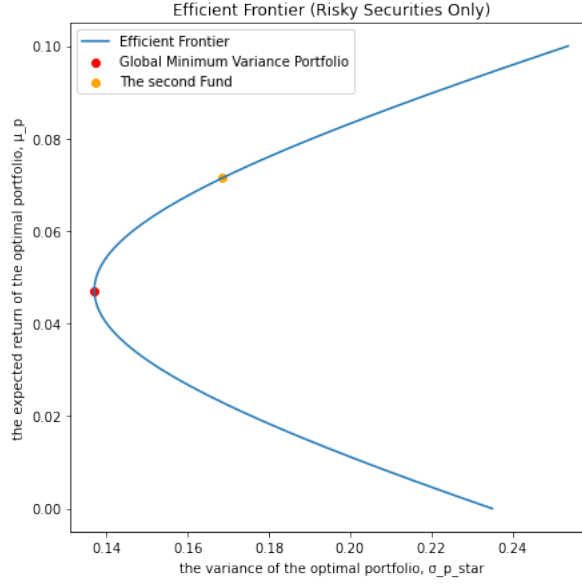
Figure 3: Two Funds Separation Theorem

security we have

$$r_f \geq 0, \quad \mathbb{E}\left[r_f\right] = r_f, \quad \text{var}\left[r_f\right] = 0, \quad \text{cov}\left[r_f, r_i\right] = 0, \text{ for all } i = 1, \ldots, n.$$

For the sake of incorporating in a risk-free asset, we have to change the weight constraint in the MVO framework to

$$w_f + \boldsymbol{w}^\top \boldsymbol{e} = 1,$$

where $w_f \in \mathbb{R}$ and $\boldsymbol{w} \in \mathbb{R}^n$ denote the (percentage) holdings in the risk-less and risky securities, respectively.

Then, we can start to construct the MVO problem with risk-free asset.

MVO Framework with risk-free Asset:

$$\min_{\boldsymbol{w}} \quad \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$$
$$\text{s.t.} \quad \boldsymbol{w}^\top \boldsymbol{\mu} + w_f r_f = \mu_p, \quad w_f + \boldsymbol{w}^\top \boldsymbol{e} = 1$$

where $\mu_p > r_f$ is desired level of portfolio return.

> **Solution:** Still, start with the Lagrangian
>
> $$L \equiv L(\boldsymbol{w}, w_f, \lambda, \gamma) = \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} + \lambda(1 - w_f - \boldsymbol{w}^\top \boldsymbol{e}) + \gamma(\mu_p - \boldsymbol{w}^\top \boldsymbol{\mu} - w_f r_f)$$
>
> By the KKT condition (Section 2.1), we have

- **Stationarity:**

$$\frac{\partial L}{\partial \boldsymbol{w}} = \boldsymbol{\Sigma}\boldsymbol{w} - \lambda \boldsymbol{e} - \gamma \boldsymbol{\mu} = \boldsymbol{0}, \tag{11}$$

$$\frac{\partial L}{\partial w_f} = \lambda + \gamma r_f = 0 \tag{12}$$

- **Primal Feasibility:**

$$\boldsymbol{w}^\top \boldsymbol{\mu} + w_f r_f = \mu_p, \tag{13}$$

$$w_f + \boldsymbol{w}^\top \boldsymbol{e} = 1 \tag{14}$$

Remark 4.25. Since there is no inequality constraints for such formulation, we omitted the **Dual Feasibility** condition and the **Complementary Slackness** condition. △

Now substitute w_f in Equation (13) using w_f in Equation (14), we have

$$\boldsymbol{w}^\top (\boldsymbol{\mu} - r_f \boldsymbol{e}) = \mu_p - r_f \tag{15}$$

From Equation (11), we get $\boldsymbol{w} = \boldsymbol{\Sigma}^{-1}(\lambda \boldsymbol{e} + \gamma \boldsymbol{\mu})$. Plug this in Equation (15), we obtain

$$(\lambda \boldsymbol{e}^\top + \gamma \boldsymbol{\mu}^\top)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f \boldsymbol{e}) = \mu_p - r_f \tag{16}$$

From Equation (12), we get $\lambda = -\gamma r_f$. Plug this in Equation (16), we obtain

$$\gamma(\boldsymbol{\mu}^\top - r_f \boldsymbol{e}^\top)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f \boldsymbol{e}) = \mu_p - r_f \tag{17}$$

Therefore, we have

$$\gamma^* = \frac{\mu_p - r_f}{(\boldsymbol{\mu} - r_f \boldsymbol{e})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f \boldsymbol{e})} \tag{18}$$

We then can obtain $\boldsymbol{w}^*$ by plugging Equation (12) into $\boldsymbol{w} = \boldsymbol{\Sigma}^{-1}(\lambda \boldsymbol{e} + \gamma \boldsymbol{\mu})$. We have

$$\boldsymbol{w}^* = \boldsymbol{\Sigma}^{-1}(-\gamma^* r_f \boldsymbol{e} + \gamma^* \boldsymbol{\mu}) = \gamma^* \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f \boldsymbol{e})$$

In all, the solution to the MVO problem with risk-free security is

$$\boldsymbol{w}^* = \gamma^* \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f \boldsymbol{e}) \tag{19}$$

where

$$\gamma^* = \frac{\mu_p - r_f}{(\boldsymbol{\mu} - r_f \boldsymbol{e})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f \boldsymbol{e})}.$$

4.3.2   Understanding the Optimal Allocation

Motivation: We want to find the relationship between μ_p and $\sigma_p = \sqrt{\boldsymbol{w}^{*\top}\boldsymbol{\Sigma}\boldsymbol{w}^*}$ for the corresponding mean-variance efficient portfolio $\boldsymbol{w}^*$.

By construction, the expected return of the optimal portfolio is

$$
\begin{aligned}
\mu_p &= \boldsymbol{w}^{*\top}\boldsymbol{\mu} + \left(1 - \boldsymbol{w}^{*\top}\boldsymbol{e}\right)r_f \\
&= r_f + \boldsymbol{w}^{*\top}(\boldsymbol{\mu} - r_f\boldsymbol{e}) \\
&= r_f + \left(\gamma^*\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f\boldsymbol{e})\right)^{\top}(\boldsymbol{\mu} - r_f\boldsymbol{e}) \\
&= r_f + \gamma^*\left(\boldsymbol{\mu} - r_f\boldsymbol{e}\right)^{\top}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - r_f\boldsymbol{e}\right)
\end{aligned}
$$

Some algebra shows that the variance of the optimal portfolio is

$$
\begin{aligned}
\sigma_p^{*2} &= \boldsymbol{w}^{*\top}\boldsymbol{\Sigma}\boldsymbol{w}^* \\
&= \left(\gamma^*\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f\boldsymbol{e})\right)^{\top}\boldsymbol{\Sigma}\left(\gamma^*\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f\boldsymbol{e})\right) \\
&= \gamma^{*2}\left(\boldsymbol{\mu} - r_f\boldsymbol{e}\right)^{\top}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - r_f\boldsymbol{e}\right)
\end{aligned}
$$

Since we only care about the case when $\mu_p > r_f$, we have

$$
\implies \sigma_p^* = \gamma^*\sqrt{\left(\boldsymbol{\mu} - r_f\boldsymbol{e}\right)^{\top}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - r_f\boldsymbol{e}\right)}
$$

In all, we have

$$
\begin{cases}
\mu_p &= r_f + \gamma^*\left(\boldsymbol{\mu} - r_f\boldsymbol{e}\right)^{\top}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - r_f\boldsymbol{e}\right) \\
\sigma_p^* &= \gamma^*\sqrt{\left(\boldsymbol{\mu} - r_f\boldsymbol{e}\right)^{\top}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - r_f\boldsymbol{e}\right)}
\end{cases}
$$

It's easy to see the above system gives a line that goes through point $(0, r_f)$ on the $(\sigma_*(\mu_p), \mu_p)$-plane:

$$
\frac{\sigma_p^* - 0}{\mu_p - r_f} = \frac{1}{\sqrt{\left(\boldsymbol{\mu} - r_f\boldsymbol{e}\right)^{\top}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - r_f\boldsymbol{e}\right)}} = \text{constant} \tag{20}
$$

Remark 4.26. This line is now **the efficient frontier** for the MVO framework with risk-free security. (The **efficient frontier** is not the hyperbola in this case).   △

We use the risk free rate specified in Equation (2), we obtain this line as presented in Figure 4 (We will show later that this is called the Capital Market Line.)

4.3.3   Tangency Portfolio

Note that, given by Equation (19),

$$
\boldsymbol{w}^* = \gamma^*\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - r_f \cdot \boldsymbol{e}\right)
$$

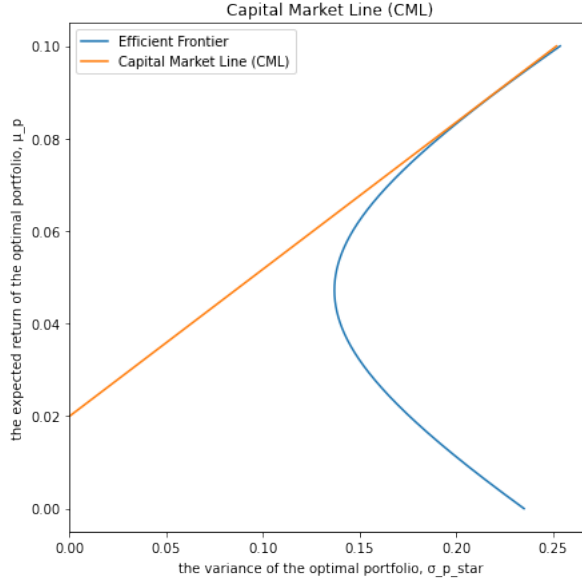**Figure 4:** Capital Market Line

implies that the weights of the risky securities of any optimal portfolio are proportional to the vector $\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-r_f\boldsymbol{e}\right)$. In other words, if we consider two optimal portfolios $\left(w_f^a,\boldsymbol{w}^a\right)$ and $\left(w_f^b,\boldsymbol{w}^b\right)$, then $\boldsymbol{w}^a=\kappa\boldsymbol{w}^b$ for some scalar $\kappa\in\mathbb{R}$.

This suggests we might want to pick an economically meaningful risky reference (benchmark) portfolio and compare all others to it.

A canonical choice is the optimal portfolio such that $w_f=0$ (i.e. an allocation only to risky securities). This portfolio is referred to as the **tangency portfolio**.

For the tangency portfolio, we must have $\boldsymbol{e}^\top\boldsymbol{w}_{Tang}=1$; that is,

$$1=\boldsymbol{e}^\top\boldsymbol{w}_{Tang}=\boldsymbol{e}^\top\gamma_{Tang}^*\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-r_f\cdot\boldsymbol{e}\right)\implies\gamma_{Tang}^*=\frac{1}{\boldsymbol{e}^\top\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-r_f\boldsymbol{e}\right)}$$

$$\implies\boldsymbol{w}_{Tang}=\gamma_{Tang}^*\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-r_f\cdot\boldsymbol{e}\right)=\frac{1}{\boldsymbol{e}^\top\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-r_f\boldsymbol{e}\right)}\cdot\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-r_f\boldsymbol{e}\right).$$

Hence

$$\boldsymbol{w}_{Tang}=\frac{1}{\boldsymbol{e}^\top\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-r_f\boldsymbol{e}\right)}\cdot\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-r_f\boldsymbol{e}\right).$$

The tangency is guaranteed by the fact that we only have one single solution for $\boldsymbol{w}_{Tang}$.

Proposition 4.27. *In a one-period economy where all investors are mean-variance optimizers, everyone shares the same mean value of returns $\boldsymbol{\mu}$ and covariance matrix of asset returns $\boldsymbol{\Sigma}$. Furthermore, investors can borrow and lend at the risk-free rate r_f. Under these conditions, one can demonstrate that in equilibrium:*
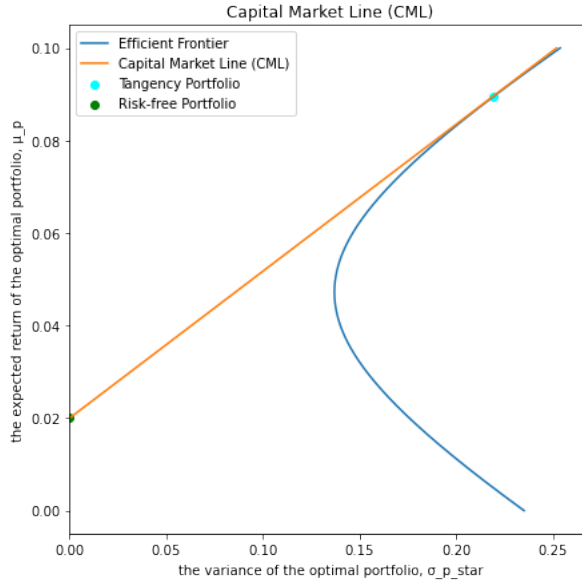
Figure 5: CML with Tangency Portfolio

(a) *the tangency portfolio consists of all securities available to investors, and*

(b) *each security in the tangency portfolio is held in proportion to its market value relative to the total market value of all securities.*

In other words, the tangency and market portfolios are the same (see Definition 4.16), i.e.

$$\boldsymbol{w}_{Tang} \equiv \boldsymbol{w}_{Market}$$

Solution: Since the efficient frontier for the MVO framework with a risk-free security is a straight line that goes through $(0, r_f)$ and $(\sigma_{Tang}, \mu_{Tang})$, all these mean-variance optimizers will invest in a portfolio consisting of the risk-free asset and the tangency portfolio corresponding to a point on this efficient frontier, as everyone shares the same mean value of returns μ and covariance matrix of asset returns Σ.

Suppose we have an asset A, and A is not in the tangency portfolio. Then, by the above logic, nobody is buying A. Consequently, the price of A will drop to the point where its expected return becomes high enough, and investors will start to include it in their portfolios, which, in this case, are the same tangency portfolio, until an equilibrium is reached. Therefore, the tangency portfolio consists of all risky securities available to investors.

Suppose the market has a total of n assets A_i, $i = 1, \ldots, n$, available to N investors in total, where each investor invests D_j, $j = 1, 2, \ldots, N$ dollars in the tangency portfolio, and $\boldsymbol{w}_{Tang} = (w_{Tang}^1, \ldots, w_{Tang}^n)$, with $\boldsymbol{w}_{Tang}^\top e = 1$. Since everyone

invests in the same risky portfolio (the tangency portfolio), for each asset A_i, its market capitalization is $C_i = \left(\sum_{j=1}^{N} D_j\right) w_{Tang}^i$. Then, the weight of asset A_i in the market portfolio, as defined in Definition 4.16 (i.e., its market value relative to the total market value of all securities), is:

$$w_M^i = \frac{C_i}{\sum_{i=1}^{n} C_i} = \frac{\left(\sum_{j=1}^{N} D_j\right) w_{Tang}^i}{\sum_{i=1}^{n} \left(\sum_{j=1}^{N} D_j\right) w_{Tang}^i} = w_{Tang}^i.$$

Therefore, each security in the tangency portfolio is held in proportion to its market value relative to the total market value of all securities.

Hence, by Definition 4.16, we see that the tangency and market portfolios are the same, i.e.,

$$\boldsymbol{w}_{Tang} \equiv \boldsymbol{w}_{Market}$$

Remark 4.28. A crucial insight from the above result is that it's not that the market portfolio is determined by the MVO framework; the market portfolio simply exists in the market. It is the tangency portfolio generated under the MVO framework, under ideal conditions, that must align with the market portfolio. △

Proposition 4.29. *The tangency / market portfolio can be determined by solving the maximal Sharpe ratio problem*

$$\max_{\boldsymbol{w}} \quad \frac{\boldsymbol{w}^\top \boldsymbol{\mu} - r_f}{\sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}}}$$
$$s.t. \quad \boldsymbol{w}^\top \boldsymbol{e} = 1$$

Remark 4.30. This formulation offers us a new perspective for the MVO framework with risk-free security where we are essentially maximizing the sharpe ratio.

△

4.3.4   The Capital Market Line (CML)

Motivation: Since we now know that under the MVO framework with risk-free security, the efficient frontier is a straight line that passes through point $(0, r_f)$ and (μ_M, σ_M), we want to represent other optimal portfolio on the line as a linear combination of these two special ones.

It immediately follows from Equation (20), we get

$$\frac{\sigma_M^* - 0}{\mu_M - r_f} = \frac{\sigma_p^* - 0}{\mu_p - r_f} = \frac{1}{\sqrt{(\boldsymbol{\mu} - r_f \boldsymbol{e})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - r_f \boldsymbol{e})}} = \text{constant}$$

where $\mu_M = w_M^\top \mu$ and $\sigma_M = \sqrt{w_M^\top \Sigma w_M}$ are the expected return and variance of the market portfolio.

Then, by simple calculation, we obtain the what's so called the capital market line (CML):

$$\mu_p = r_f + \sigma_p^* \left[\frac{\mu_M - r_f}{\sigma_M} \right] \tag{21}$$

Remark 4.31.

1. The bracketed term in the CML, $\left[\frac{\mu_M - r_f}{\sigma_M} \right]$, is referred to as **the market price of risk**.

   Its **economic meaning** can be interpret as follows:

   - The numerator $\mu_M - r_f$ is the expected return from investing in the market above the risk-free return. It is a measure of the reward for holding the risky market portfolio rather than the risk-free security.
   - The denominator σ_M is the market risk of the market portfolio.

   Thus the slope of the CML, $\left[\frac{\mu_M - r_f}{\sigma_M} \right]$, measures the reward per unit of market risk, determines the additional return needed to compensate for a unit change in risk.

   In other words, the CML says that the expected return on a portfolio is equal to the risk-free rate (r_f) plus a risk premium $\left(\sigma_p^* \left[\frac{\mu_M - r_f}{\sigma_M} \right] \right)$, where the risk premium is equal to the market price of risk, $\left[\frac{\mu_M - r_f}{\sigma_M} \right]$, (as measured by the reward per unit of market risk) times the quantity of risk for the portfolio, σ_p^*, (as measured by the standard deviation of the portfolio).

2. The CML is the two-fund separation theorem in the case of multiple risky securities and one risk-less security. Indeed, as we can see, any efficient portfolio can be attained by a linear combination of the "risk-free portfolio" and market portfolio.

   $\triangle$

The following figure illustrates the CML. Notice the CML is a tangent line to the efficient frontier (with only risk securities) at the market/tangency portfolio.

Remark 4.32.

1. An investor will select a portfolio on the CML that represents a combination of borrowing or lending at the risk-free rate and the market portfolio.

2. Portfolios to the left of the market portfolio/tangency portfolio represent combinations of risky securities and the risk-free security.

3. Portfolios to the right of the market portfolio include purchases of risky securities made with funds borrowed at the risk-free rate. Such a portfolio is called a leveraged portfolio because it involves the use of borrowed funds.
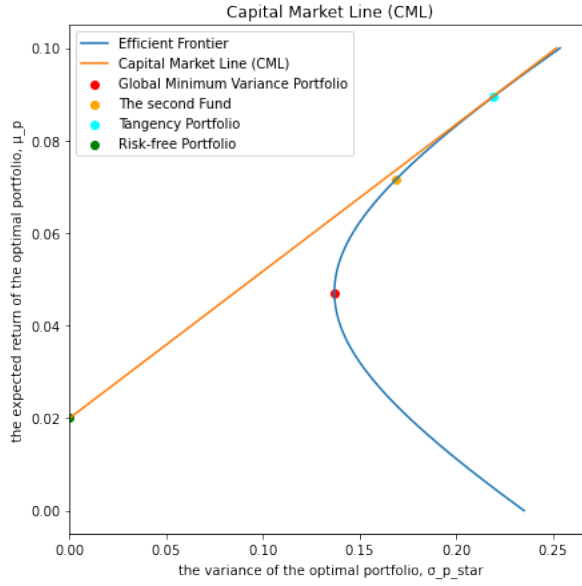
   $\triangle$

Figure 6: MVO Summary

Finally, we present the solution to the Risk Aversion Formulation of MVO that has constraints on portfolio weights here at the end, as it is quite popular in some contexts due to its simple solution.

Problem 4.33. (The Risk Aversion Formulation) The risk aversion formulation is stated as

$$\max_{\boldsymbol{w}} \quad \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\lambda}{2} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$$
$$\text{s.t.} \quad \lambda = \lambda_0$$

where $\lambda \geq 0$ can be interpreted as the the risk aversion coefficient specified by the investor.

Solution: Set $L(\boldsymbol{w}) = \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\lambda_0}{2} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$. We get

$$\frac{\partial L}{\partial \boldsymbol{w}} = \boldsymbol{\mu} - \lambda_0 \boldsymbol{\Sigma} \boldsymbol{w} \implies \boldsymbol{w}^* = (\lambda_0 \boldsymbol{\Sigma})^{-1} \boldsymbol{\mu}$$

Hence,

$$\boldsymbol{w}^* = \lambda_0^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$$

4.4   Drawbacks of MVO

Motivation: We want to know how to apply the MVO framework properly in practice as well as its advantages and disadvantages.

We will use the data presented in Section 3 to better illustrate our points.

Suppose the expected returns for all countries for the next five years are 5%. Then, according to the classical MVO framework, we obtain our portfolio weights. However, since expected returns are difficult to predict, let's consider what would happen if our views were to shift slightly. Suppose the expected return for Germany changes to 7%, while those for France and the UK both change to 4%. In this case, we can calculate the optimal portfolio weights using the classical MVO framework with the updated expected returns. Then, finally, we can compare the two optimal portfolio weights as presented in the Figure 7 below:
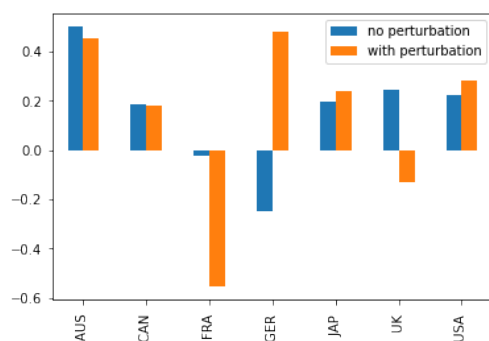

Figure 7: Optimal portfolio weights with and without perturbation

We see that the weights change drastically with small perturbations, which means the classical MVO framework is not robust. On the other hand, we had significantly shorted Germany and gone long on the UK before. However, after adjusting the expected returns, we ended up going heavily long on Germany, shorting the UK, and heavily shorting France. The changes make the portfolio weights yielded by the MVO framework hard to interpret. Moreover, since we have to make estimations for the expected returns, we have no means to tackle the estimation errors, let alone the fact that it is impossible for investors to have meaningful, accurate insights for thousands of stocks currently trading in the market. Further, the classical MVO framework restricts the views to be only on the expected returns of the assets. If we want to incorporate relative views, such as stock B outperforming stock A by 5%, we can't do that.

Therefore, we can conclude four drawbacks of the classical MVO framework:

1. The classical MVO can result in counter-intuitive portfolios.

2. Significant impact from estimation error in expected returns and covariances.

3. The investor is required to provide estimates of expected returns and covariances of all assets.

4. Hard to incorporate relative forecasts: "B will outperform A by 0.5% over the next month".

5   The Capital Asset Pricing Model – CAPM

Motivation: We want to analyze how the expected rate of return of an individual asset $\mathbb{E}[r_i]$ is in relation to its individual risk β_i with respect to the market equilibrium.

Assumption 5.1. *(Core Assumptions)*

  (a) *All investors use mean-variance analysis to select a portfolio.*

  (b) *All investors have homogeneous (same) believes about the future return, variance and covariance of assets.*

  (c) *There is a unique risk-free rate of borrowing and lending available for all investors.*

  (d) *There are no transaction costs.*

Under these assumptions, as discussed in Proposition 4.27 and Section 4.3.4 that an investor will select a portfolio on the CML that represents a combination of borrowing or lending at the risk-free rate and the market portfolio..

5.1   The Pricing Formula

Theorem 5.2. *(The capital asset pricing model) If the market portfolio M is mean-variance efficient, the expected return $\mathbb{E}[r_i]$ of an asset i satisfies*

$$\mathbb{E}\left[r_i\right] - r_f = \beta_i \left(\mathbb{E}\left[r_M\right] - r_f\right)$$

where

  • $\beta_i = \frac{Cov(r_i, r_M)}{\sigma_M^2}$ *is often called the beta of asset i.*

  • σ_M^2 *represents the variance of the market portfolio.*

Solution: Consider the portfolios spanned by individual asset i and the market portfolio; that is $r_\alpha = \alpha r_i + (1-\alpha) r_M$ where $\alpha \in \mathbb{R}$. The expected rate of return of this portfolio is then

$$\mathbb{E}[r_\alpha] = \alpha \mathbb{E}[r_i] + (1-\alpha)\mathbb{E}[r_M]$$

The variance of the return is

$$\sigma_\alpha^2 = \alpha^2 \sigma_i^2 + 2\alpha(1-\alpha)Cov\left(r_i, r_M\right) + (1-\alpha)^2 \sigma_M^2$$

As α varies, the values of $(\sigma_\alpha, \mathbb{E}[r_\alpha])$ trace out a curve on the $(\sigma, \mathbb{E}[r])$ diagram, as shown in the Figure 8 below. In particular, $\alpha = 0$ corresponds to the market portfolio M. The curve cannot cross the capital market line, as this is the efficient frontier. Hence at $\alpha = 0$, the curve must be tangent to the capital market line at M,

and thus having the same slope at M. Since from Section 4.3.4, we know the slope of the Capital Market Line is $\frac{\mathbb{E}[r_M]-r_f}{\sigma_M^2}$, we must have

$$\frac{\mathbb{E}[r_M]-r_f}{\sigma_M^2} = \frac{\partial \mathbb{E}[r_\alpha]}{\partial \sigma_\alpha}\bigg|_{\alpha=0}$$

$$= \frac{\partial \mathbb{E}[r_\alpha]}{\partial \alpha} \frac{\partial \alpha}{\partial \sigma_\alpha}\bigg|_{\alpha=0}$$

$$= \frac{\partial \mathbb{E}[r_\alpha]}{\partial \alpha} \left(\frac{\partial \sigma_\alpha}{\partial \alpha}\right)^{-1}\bigg|_{\alpha=0}$$

$$= (\mathbb{E}[r_i] - \mathbb{E}[r_M]) \cdot \frac{\sigma_\alpha}{\alpha \sigma_i^2 + (1-2\alpha)Cov\,(r_i, r_M) + (\alpha-1)\sigma_M^2}\bigg|_{\alpha=0}$$

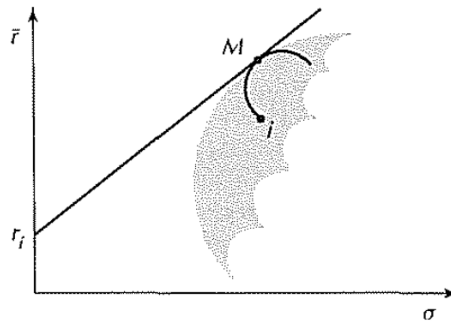$$= \frac{\mathbb{E}[r_i] - \mathbb{E}[r_M]}{Cov\,(r_i, r_M) - \sigma_M^2}$$

Hence,

$$\frac{\mathbb{E}[r_i] - \mathbb{E}[r_M]}{Cov\,(r_i, r_M) - \sigma_M^2} = \frac{\mathbb{E}[r_M]-r_f}{\sigma_M^2}$$

Solve for $\mathbb{E}[r_i]$, to obtain the final result

$$\mathbb{E}[r_i] = \frac{Cov\,(r_i, r_M)}{\sigma_M^2}\left(\mathbb{E}[r_M] - r_f\right) + r_f$$

Set $\beta_i = \frac{Cov(r_i, r_M)}{\sigma_M^2}$, we obtain

$$\mathbb{E}[r_i] = \beta_i\left(\mathbb{E}[r_M] - r_f\right) + r_f$$

Figure 8: Portfolio curve: The family of portfolios traces out n curve on the diagram This curve cannot cross the **capital market line**, and hence must be tangent to that line

Remark 5.3.

1. The CAPM model concludes that the expected excess rate of return of an asset $\mathbb{E}[r_i] - r_f$ is proportional to the expected excess rate of return of the market portfolio $\mathbb{E}[r_M] - r_f$ and the market beta $\beta_i = \frac{Cov(r_i, r_M)}{\sigma_M^2}$.

2. The CAPM changes our concept of the risk of an asset from that of a σ that of β. It is still true that, overall, we measure the risk of a portfolio in terms of σ, but this does not translate into a concern for the σ's of **individual assets**. For those, the propel measure is their β's. That is,

$$\text{risk of a portfolio} \quad \longrightarrow \quad \sigma$$
$$\text{risk of an individual asset} \quad \longrightarrow \quad \beta$$

$\triangle$

5.2 Understanding the CAPM Model

To gain insight into this result, let us consider some extreme cases:

1. Suppose that the asset is completely uncorrelated with the market; that is, $\beta_i = 0$. Then, according to the CAPM, we have $\mathbb{E}[r_i] = r_f$, which means that even if the asset is very risky (with large σ), the expected rate of return will be the risk-free rate – there is no premium for risk.

    This happens because the risk linked to an asset that doesn't move in sync with the market can be spread out or diversified. Imagine having many such assets, each one not related to the others or the market. If we bought little bits of all these assets, the overall risk or variance would be small. Because the combined return from all these assets would also have a small variance, the expected rate of return should be near r_f, which has no variance at all.

2. An even more extreme case is when an asset has a negative value of β_i. In this situation, $\mathbb{E}[r_i] < r_f$. This means that even if the asset has a high risk (indicated by its σ), its expected rate of return should be even lower than the risk-free rate.

    This happens because such an asset actually decreases the total risk of the portfolio when it's combined with the market. So, investors are ready to accept a lower expected return due to this risk-reducing advantage. In a sense, these types of assets act like insurance. They perform well when everything else is performing poorly.

5.3 CAPM in Matrix-vector Form

Denote by $\boldsymbol{w}_M = (w_{M,1}, \ldots, w_{M,n})$ the normalized market capitalization (or benchmark) weights such that $w_M^\top e = 1$. Then the return on the market can be expressed as

$$r_M = \sum_{i=1}^{n} w_{M,i} r_i$$

By CAPM, the expected excess return on security i becomes

$$\pi_i = \mathbb{E}[r_i] - r_f = = \beta_i \left(\mathbb{E}\left[r_M\right] - r_f \right)$$
$$= \frac{Cov\left(r_i, r_M\right)}{\sigma_M^2} \left(\mathbb{E}\left[r_M\right] - r_f \right)$$
$$= \frac{\mathbb{E}\left[r_M\right] - r_f}{\sigma_M^2} \sum_{j=1}^{n} \text{cov}\left(r_i, r_j\right) w_{M,j}$$

Defining $\delta = \frac{\mathbb{E}[r_M] - r_f}{\sigma_M^2}$, we write CAPM in the matrix-vector form

$$\boldsymbol{\pi} = \delta \boldsymbol{\Sigma} \boldsymbol{w}_M$$

Remark 5.4. π is referred to as the "market implied expected returns" or "equilibrium implied expected returns". $\triangle$

5.4   Security Market Line

The Security Market Line is nothing more than a visual representation of the CAPM formula on the $(\beta, \mathbb{E}[r])$-plane or $(Cov(r, r_M), \mathbb{E}[r])$-plane.
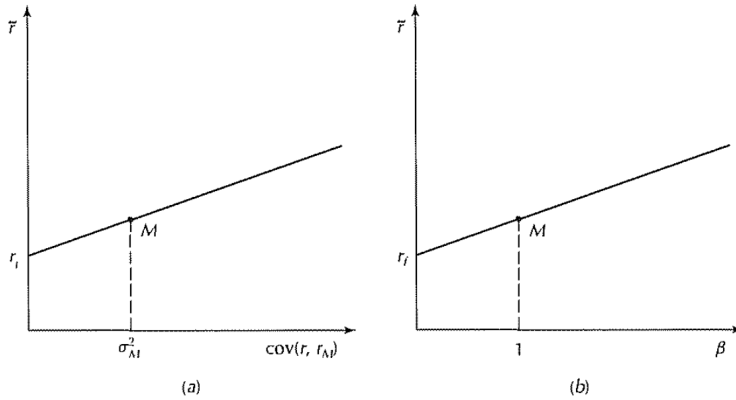
Figure 9: Security Market Line: The expected rate of return increases linearly as the covariance with the market increases or, equivalently, as β increases

Remark 5.5. The Security Market Line highlights the essence of the CAPM formula: **Under the equilibrium conditions assumed by the CAPM, any asset should fall on the security market line.** $\triangle$

5.5   Systematic and Non-systematic Risks

Proposition 5.6. *Suppose we write the return* (r_i) *of asset* i *as*

$$r_i = r_f + \beta_i \left(r_M - r_f \right) + \varepsilon_i.$$

where r_f is a constant, r_i, r_M are random variables, $\beta_i = \frac{Cov(r_i, r_M)}{\sigma_M^2}$,and ε_i is a random variable to indicate the uncertainty in the return.

Then, by CAPM, we must have

(a) $\mathbb{E}[\varepsilon_i] = 0$.

(b) $Cov(\varepsilon_i, \sigma_M) = 0$.

(c) The variance of an asset is:
$$\sigma_i^2 = \beta_i^2 \sigma_M^2 + Var(\varepsilon_i)$$

Solution:

(a) This one is trivial. We can just take the expectation on both sides, then we have our result.

(b) Since $\beta_i = \frac{Cov(r_i, r_M)}{\sigma_M^2}$ can be viewed as a constant, we have

$$
\begin{aligned}
Cov(r_i, r_M) &= Cov(r_f + \beta_i(r_M - r_f) + \varepsilon_i, r_M) \\
&= Cov(\beta_i r_M + \varepsilon_i, r_M) \\
&= Cov(\beta_i r_M, r_M) + Cov(\varepsilon_i, r_M) \\
&= \beta_i \sigma_M^2 + Cov(\varepsilon_i, r_M) \\
&= \frac{Cov(r_i, r_M)}{\sigma_M^2} \sigma_M^2 + Cov(\varepsilon_i, r_M) \\
&= Cov(r_i, r_M) + Cov(\varepsilon_i, r_M)
\end{aligned}
$$

Therefore, we must have
$$Cov(\varepsilon_i, r_M) = 0.$$

(c) We have

$$
\begin{aligned}
Var(r_i) = \sigma_i^2 &= Var(r_f + \beta_i(r_M - r_f) + \varepsilon_i) \\
&= Var(\beta_i r_M + \varepsilon_i) \\
&= \beta_i^2 \sigma_M^2 + 2\beta_i Cov(\varepsilon_i, r_M) + Var(\varepsilon_i) \\
Cov(\varepsilon_i, r_M) = 0 \implies &= \beta_i^2 \sigma_M^2 + Var(\varepsilon_i)
\end{aligned}
$$

The Key Insights are:

For the variance of the rate of return r_i of an individual asset as represented in the form

$$\sigma_i^2 = \beta_i^2 \sigma_M^2 + Var(\varepsilon_i) \tag{22}$$

- The first part $\beta_i^2 \sigma_i^2$ is called **systematic risk**. This is the risk associated with the market as a whole. This risk cannot be reduced by diversification because every asset with nonzero beta contains this risk.

- The second part, $Var(\varepsilon_i)$, is termed the **non-systematic**, **idiosyncratic**, or **specific risk**. This risk is uncorrelated with the market and can be reduced by diversification.

It is the systematic (or non-diversifiable) risk, measured by beta, that is most important, since it directly combines with the systematic risk of other assets.

Remark 5.7. This result emphasizes that the risk of an asset is a function of its covariance with the market or, equivalently, a function of its beta. △

Connections with the Capital Market Line:

Consider an asset r_i on the capital market line with a value of β_i. Then, by the equation of CML, we have

$$\mathbb{E}[r_i] = r_f + \sigma_i \left[\frac{\mathbb{E}[r_M] - r_f}{\sigma_M} \right] = r_f + \frac{\sigma_i}{\sigma_M} (\mathbb{E}[r_M] - r_f)$$

$$\implies \beta = \frac{\sigma_i}{\sigma_M} \implies \sigma_i = \beta \sigma_M$$

Therefore, the standard deviation of this asset is $\sigma_i = \beta \sigma_M$. Comparing this with Equation (22), we obtain that assets on the capital market line have only systematic risk; there is no nonsystematic risk. This asset has an expected rate of return equal to $\mathbb{E}[r_i] = r_f + \beta (\mathbb{E}[r_M] - r_f)$.

Now consider a whole group of other assets, all with the same value of β. According to CAPM, these all have the same expected rate of return, equal to $\mathbb{E}[r_i]$. However, if these assets carry nonsystematic risk, they will not fall on the capital market line. Indeed, as the nonsystematic risk increases, the points on the $(\sigma, \mathbb{E}[r])$-plane representing these assets drift to the right, as shown in Figure 10 The hoizontal distance of a point from the capital market line is therefore a measure of the nonsystematic risk.
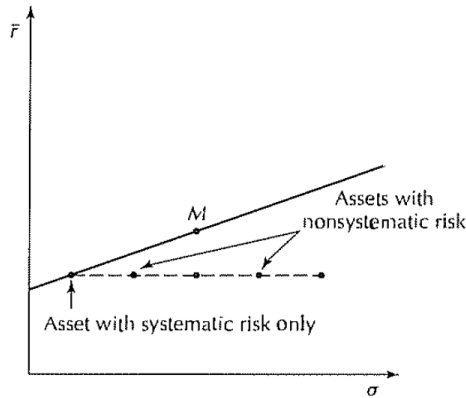
Figure 10: Systematic and nonsystematic risk An asset on the **capital market line** has only systematic risk. Assets with nonsystematic risk fall to the right of the **capital market line**

5.6   Sharpe Ratio

In the $(\sigma, \mathbb{E}[r])$-plane, we can easily demonstrate the Sharpe ratio of a feasible portfolio. From Figure 11, we can easily understand the intuition behind why the market portfolio
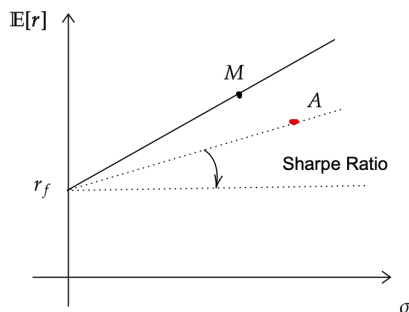

Figure 11: Sharpe Ratio

is the one that maximizes the Sharpe ratio, as specified in Proposition 4.29.

6 Arbitrage Pricing Theory – Factor Model

Motivation: We want to generalize the CAPM model. Therefore, we can interpret the APT model as a single-period static model generalized from CAPM.

6.1 Formulation

Assumption 6.1. *(Core Assumptions)*

1. *The average investor exhibits risk aversion (Definition 4.12 and non-satiation (Definition 4.13) preferences.*

2. *Market is efficient with limited opportunity for arbitrage.*

3. *Perfect capital markets*

4. *Infinite number of assets*

5. *Risk factors are indicative of systematic risks that cannot be diversified away and thus impact all financial assets, to some degree. Thus, these factors must be: Non-specific to any individual firm or industry Compensated by the market via a risk premium A random variable*

APT model Framework: Assume that the assumptions hold.

Then, risky asset returns are said to follow a **factor intensity structure** if the return of each security $i(i = 1, \ldots, n)$ at time $t(t = 1, \ldots, T)$ can be expressed as a multifactor model:

$$r_{it} = \alpha_i + \beta_{i1}f_{1t} + \beta_{i2}f_{2t} + \cdots + \beta_{ik}f_{kt} + \varepsilon_{it} = \alpha_i + \sum_{j=1}^{k} \beta_{ij}f_{jt} + \varepsilon_{it}$$

where $k \ll n < T$, $\mathbb{E}(\varepsilon_{it}) = \mathbb{E}(f_{jt}) = \mathbb{E}(\varepsilon_{it}f_{jt'}) = \mathbb{E}(\varepsilon_{it}\varepsilon_{lt'}) = 0$ $(i \neq j, i \neq l,$ and for all $t, t')$ also $\mathbb{E}(\varepsilon_{it}^2) \equiv \sigma_i^2$.

- α_i is a constant for asset i,

- f_{jt} is the jth factor/systematic factor/risk factor at time t and can be viewed as a random variable,

- β_{ij} factor beta/loading or sensitivity of the ith security to the jth factor,

- ε_{it} is nonsystematic (idiosyncratic) return of the ith security at time t.

Remark 6.2.

1. We require $k \ll n$ because we don't want our matrix for betas to be rank deficient, which is essential for solving linear regression problems accurately and reliably.

2. $\mathbb{E}[r_{it}] = \alpha_i$.

3. $\mathbb{E}\left(\varepsilon_{it}\varepsilon_{lt'}\right) = 0$ is often not satisfied in practice because it is impossible to find all the factors that would perfectly express the correlation between assets.

4. Setting $k = 1$ and $f_{it} = r_{Mt} - r_{ft}$, where r_{Mt} and r_{ft} represent the return of the market portfolio and the risk-free rate at time t respectively, yields the CAPM model.

5. A disadvantage of APT is that the selection and the number of factors to use in the model is ambiguous. Most academics use three to five factors to model returns, but the factors selected have not been empirically robust. In many instances the CAPM, as a model to estimate expected returns, has empirically outperformed the more advanced APT.

$\triangle$

6.2 APT in Matrix Form

Generalizing further and aligning with the notation often used in the linear regression context, we can re-formulate the APT model as

$$\boldsymbol{r} = \mathbf{X}\boldsymbol{f} + \boldsymbol{\varepsilon}, \quad \mathbb{E}[\boldsymbol{\varepsilon}] = 0, \quad \mathbb{V}[\boldsymbol{\varepsilon}] = \mathbf{D}$$

where

- $\boldsymbol{r} \in \mathbb{R}^n$ is an n-dimensional random vector containing the crosssection of **returns in excess of the risk-free rate** over some time interval $[t, t+1]$,

- $\mathbf{X} \in \mathbb{R}^{n \times k}$ is a (non-random) $n \times k$ matrix for factor loadings that is known before time t,

- $\boldsymbol{f} \in \mathbb{R}^k$ denotes a k-dimensional random vector process for latent factors which cannot be observed directly,

- $k \ll n$,

- $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ is assumed to follow a mean-zero distribution with diagonal variance-covariance matrix given by $\mathbf{D} = diag\left(\sigma_1^2, \ldots, \sigma_n^2\right)$ with all $\sigma_i^2 > 0$.

Remark 6.3. Information about the $\boldsymbol{f}$-process must be obtained via statistical inference. $\triangle$

Proposition 6.4. *(APT Useful Technique) **In general, it is useful** to assume that the $\boldsymbol{f}$-process is Gaussian and has finite first and second moments given by*

$$\mathbb{E}[\boldsymbol{f}] = \boldsymbol{\mu}_f, \text{ and } \mathbb{V}[\boldsymbol{f}] = \mathbf{F}$$

which yields

$$\mathbb{E}[\boldsymbol{r}] = \mathbf{X}\boldsymbol{\mu}_f, \text{ and } \mathbb{V}[\boldsymbol{r}] = \mathbf{D} + \mathbf{X}\mathbf{F}\mathbf{X}^\top \tag{23}$$

i.e.,

$$\boldsymbol{r} \sim \mathcal{N}\left(\mathbf{X}\boldsymbol{\mu}_f, \mathbf{D} + \mathbf{X}\mathbf{F}\mathbf{X}^\top\right)$$

Remark 6.5. The elements of $\boldsymbol{\mu}_f$ are called factor risk premia. $\triangle$

6.3   APT model as a way to estimate covariance matrix

Suppose the returns satisfies the APT model, then we have

$$r = \mathbf{X}f + \varepsilon, \quad \mathbb{E}[\varepsilon] = 0, \quad \mathbb{V}[\varepsilon] = \mathbf{D} = diag\left(\sigma_1^2, \ldots, \sigma_n^2\right) \text{ with all } \sigma_i^2 > 0.$$

Setting f process to be Gaussian ww get

$$r \sim \mathcal{N}\left(\mathbf{X}\boldsymbol{\mu}_f, \mathbf{D} + \mathbf{X}\mathbf{F}\mathbf{X}^\top\right)$$

The MVO yields:

$$
\begin{aligned}
\boldsymbol{h}^* &= \arg\max_{\boldsymbol{h}} \mathbb{E}[\boldsymbol{h}^\top r] - \frac{\lambda}{2}\mathbf{V}[\boldsymbol{h}^\top r] \\
&= \arg\max_{\boldsymbol{h}} \boldsymbol{h}^\top \mathbf{X}\boldsymbol{\mu}_f - \frac{\lambda}{2}\boldsymbol{h}^\top (\mathbf{X}\mathbf{F}\mathbf{X}^\top + \mathbf{D})\boldsymbol{h} \\
&= \arg\max_{\boldsymbol{h}} \boldsymbol{h}^\top \mathbf{X}\boldsymbol{\mu}_f - \frac{\lambda}{2}\boldsymbol{h}^\top \mathbf{X}\mathbf{F}\mathbf{X}^\top \boldsymbol{h} - \frac{\lambda}{2}\boldsymbol{h}^\top \mathbf{D}\boldsymbol{h} \\
\boldsymbol{q} := \mathbf{X}^\top \boldsymbol{h} \Longrightarrow &= \arg\max_{\boldsymbol{h}} \boldsymbol{q}^\top \boldsymbol{\mu}_f - \frac{\lambda}{2}\boldsymbol{q}^\top \mathbf{F}\boldsymbol{q} - \frac{\lambda}{2}\boldsymbol{h}^\top \mathbf{D}\boldsymbol{h}
\end{aligned}
$$

and where $\lambda > 0$ is the Arrow-Pratt constant absolute risk aversion. We call the third term, $\boldsymbol{h}^\top \mathbf{D}\boldsymbol{h}$ as "idiosyncratic variance". To proceed, we want to write this term as a function of $\boldsymbol{q}$ as well.

Proposition 6.6. *An optimal portfolio h^* specified above must minimise idiosyncratic variance $h^\top \mathbf{D} h$ among all portfolios with the same exposures $q = \mathbf{X}^\top h$.*

Theorem 6.7. *The risk/alpha exposures q^* and the optimal holdings h^* are*

$$
\begin{aligned}
\boldsymbol{q}^* &= \lambda^{-1}\left[\mathbf{F} + \left[\mathbf{X}^\top \mathbf{D}^{-1}\mathbf{X}\right]^+\right]^{-1}\boldsymbol{\mu}_f \\
\boldsymbol{h}^* &= \lambda^{-1}\mathbf{D}^{-\frac{1}{2}}\left(\mathbf{X}^\top \mathbf{D}^{-\frac{1}{2}}\right)^+\left[\mathbf{F} + \left(\mathbf{X}^\top \mathbf{D}^{-1}\mathbf{X}\right)^+\right]^{-1}\boldsymbol{\mu}_f
\end{aligned}
$$

where $\mathbf{X}^+$ stands for the pseudo-inverse of matrix $\mathbf{X}$.

Solution: Since $\mathbf{X}^\top h = q$ and $\mathbf{D}$ is diagonal, we have

$$\boldsymbol{h}^*(\boldsymbol{q}) = \arg\min_{\boldsymbol{h}} \boldsymbol{h}^\top \mathbf{D}\boldsymbol{h}$$

$$\boldsymbol{\eta} := \mathbf{D}^{\frac{1}{2}}\boldsymbol{h} \Longrightarrow \boldsymbol{\eta}^* = \arg\min_{\boldsymbol{\eta}} \|\boldsymbol{\eta}\|^2 \text{ subject to } \mathbf{X}^\top \mathbf{D}^{-\frac{1}{2}}\boldsymbol{\eta} = \boldsymbol{q}$$

This can be viewed as an ordinary least squares problem; thus, we have

$$\boldsymbol{\eta}^* = \left(\mathbf{X}^\top \mathbf{D}^{-\frac{1}{2}}\right)^+ \boldsymbol{q} \quad \Longrightarrow \quad \boldsymbol{h}^*(\boldsymbol{q}) = \mathbf{D}^{-\frac{1}{2}}\left(\mathbf{X}^\top \mathbf{D}^{-\frac{1}{2}}\right)^+ \boldsymbol{q}$$

Let $V(q) = h^*(q)^\top \mathbf{D} h^*(q)$ be the minimum idiosyncratic variance (still subject to $\mathbf{X}^\top h = q$). Using Proposition 6.6, the mean-variance objective can then be written entirely in terms of q:

$$q^* = \arg\max_q \left\{ q \cdot \mu_f - \frac{\lambda}{2} q^\top \mathbf{F} q - \frac{\lambda}{2} V(q) \right\}$$

Additionally, we have

$$V(q) = h^*(q)^\top \mathbf{D} h^*(q)$$
$$= \left[\mathbf{D}^{-\frac{1}{2}} \left(\mathbf{X}^\top \mathbf{D}^{-\frac{1}{2}} \right)^+ q \right]^\top \mathbf{D} \left[\mathbf{D}^{-\frac{1}{2}} \left(\mathbf{X}^\top \mathbf{D}^{-\frac{1}{2}} \right)^+ q \right]$$
$$= q^\top \left[\mathbf{X}^\top \mathbf{D}^{-1} \mathbf{X} \right]^+ q$$

and therefore:

$$q^* = \arg\max_q \left\{ q \cdot \mu_f - \frac{\lambda}{2} q^\top \mathbf{F} q - \frac{\lambda}{2} V(q) \right\}$$
$$= \arg\max_q \left\{ q \cdot \mu_f - \frac{\lambda}{2} q^\top \mathbf{F} q - \frac{\lambda}{2} q^\top \left[\mathbf{X}^\top \mathbf{D}^{-1} \mathbf{X} \right]^+ q \right\}$$
$$= \arg\max_q \left\{ q \cdot \mu_f - \frac{\lambda}{2} q^\top \left[\mathbf{F} + \left[\mathbf{X}^\top \mathbf{D}^{-1} \mathbf{X} \right]^+ \right]^{-1} q \right\}$$
$$= \lambda^{-1} \left[\mathbf{F} + \left[\mathbf{X}^\top \mathbf{D}^{-1} \mathbf{X} \right]^+ \right]^{-1} \mu_f$$

and the optimal holdings h^* are

$$h^*(q) = \mathbf{D}^{-\frac{1}{2}} \left(\mathbf{X}^\top \mathbf{D}^{-\frac{1}{2}} \right)^+ q$$
$$= \lambda^{-1} \mathbf{D}^{-\frac{1}{2}} \left(\mathbf{X}^\top \mathbf{D}^{-\frac{1}{2}} \right)^+ \left[\mathbf{F} + \left(\mathbf{X}^\top \mathbf{D}^{-1} \mathbf{X} \right)^+ \right]^{-1} \mu_f$$

Theorem 6.8. *The computational complexity of the above method is $O(n)$ if $k << n$.*

Solution: The only complicated thing involved in the above method is to compute the pseudo-inverse of $\mathbf{X}^\top \mathbf{D}^{-\frac{1}{2}} \in \mathbb{R}^{n \times k}$ for which we just have to compute the SVD of it where (Golub & Van Loan 2012) showed the computational complexity is about:

$$6nk^2 + 20k^3 \sim O(n)$$

since $k \ll n$.

Remark 6.9.

1. We do not assume X to be of full rank. Thus, this method is a good tool for handling collinearities in X.

2. This method is proposed in Ritter(2016). I just present it here as a reference.

$\triangle$

7   Classical Black-Litterman Framework

In the standard Mean-Variance Optimization (MVO) approach, an investor needs to estimate expected returns and covariances for all securities in their investment universe. With the sheer number of securities available today, this task is daunting. Often, investors focus on a specific area to earn better returns, which can result in a lot of estimation errors. Moreover, if an investor has more confidence in some estimates than others, it makes sense to treat these inputs differently when building portfolios. But, you can't do this in the traditional MVO approach. It's important to note that including external insights into formal models is crucial in practice. These insights can be valuable inputs, and portfolio managers might not want to hand over control to an automated system or 'black box'. This is where the Black-Litterman model comes in. It's a modified MVO that allows external insights, like a portfolio manager's judgment, to be included in formal models using Bayesian techniques. It addresses the issue of estimation errors, providing a way for portfolio managers to have better control over a quantitative framework

7.1   General Assumptions

Assumption 7.1. *(Base Assumption) Unless the investor has a specific view of the security, the expected return of a security should be consistent with market equilibrium.*

Remark 7.2. In other words, an unconstrained investor who does not have any views on the market should hold the market. (Gain beta if you can't earn alpha.)   △

Assumption 7.3. *(Core Assumptions)*

(a) *Securities:*
$$r \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

    Security returns form a multivariate normally distributed random vector.

(b) *Market equilibrium (CAPM prior):*
$$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\pi}, \tau\boldsymbol{\Sigma})$$

    The vector of expected return, $\boldsymbol{\mu}$, is itself a multivariate normally distributed random vector.

(c) *Investor's views:*
$$\mathbf{P}\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{q}, \boldsymbol{\Omega})$$

where

- $\boldsymbol{\mu} \in \mathbb{R}^n$ *is the vector of expected return,*

- $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ *is the covarince matrix of returns,*

- $\boldsymbol{\pi} \in \mathbb{R}^n$ *is the vector of CAPM equilibrium returns (prior),*

- $0 < \tau < 1, \tau \in \mathbb{R}$ *is a parameter to express the uncertainty in the CAPM equilibrium,*

- $q \in \mathbb{R}^k$ *is the vector of the expected returns of investor views (forecasts /"alphas"),*

- $\mathbf{P} \in \mathbb{R}^{k \times n}$ *the matrix of investor's views, and sometimes referred to as the "picking matrix" as it "picks out" the securities that the investor has views about.*

- $\mathbf{\Omega} \in \mathbb{R}^{k \times k}$ *is the covariance matrix of the views, representing the uncertainty in the investor views. (the "confidence")*

Remark 7.4.

1. Assumptions (a) and (b) encapsulate the idea that the expected returns of the securities deviate from the perceived market equilibrium.

2. Naturally, both the expected return vector and the vector of market equilibrium returns are not observable and have to be estimated.

$\triangle$

Incorporate Investor's Views

Example 7.5. (Investor's View) Let us assume we have four assets and three views ($n = 4, k = 3$):

$\mathbb{E}[r_1] = 3\% \pm 3.9\%$ with 95% confidence $(1.96 \times 2\% = 3.92\%) \Longleftarrow \hat{se}_1$ or $\sigma_1 = 2\%$,

$\mathbb{E}[r_2] = 4\% \pm 5.9\%$ with 95% confidence $(1.96 \times 3\% = 5.88\%) \Longleftarrow \hat{se}_2$ or $\sigma_2 = 3\%$,

$\mathbb{E}[r_4] - \mathbb{E}[r_3] = 1\% \pm 2\%$ with 95% confidence $(1.96 \times 1\% = 1.96\%) \Longleftarrow \hat{se}_3$ or $\sigma_3 = 1\%$.

We express the views as

$$q + \varepsilon = \mathbf{P}\mu \quad \text{where} \quad \varepsilon \sim \mathcal{N}(0, \mathbf{\Omega})$$

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}, q = \begin{bmatrix} 3\% \\ 4\% \\ 1\% \end{bmatrix}, \quad \mathbf{\Omega} = \begin{bmatrix} 2\%^2 & 0 & 0 \\ 0 & 3\%^2 & 0 \\ 0 & 0 & 1\%^2 \end{bmatrix}$$

Remark 7.6.

1. **(More detail)** That is,

$$q + \varepsilon = \begin{bmatrix} 3\% \\ 4\% \\ 1\% \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \mathbb{E}[r_1] \\ \mathbb{E}[r_2] \\ \mathbb{E}[r_3] \\ \mathbb{E}[r_4] \end{bmatrix} = \begin{bmatrix} \mathbb{E}[r_1] \\ \mathbb{E}[r_2] \\ \mathbb{E}[r_4] - \mathbb{E}[r_3] \end{bmatrix}$$

where

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \sim \mathcal{N}(0, \mathbf{\Omega}), \quad \mathbf{\Omega} = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix} = \begin{bmatrix} 2\%^2 & 0 & 0 \\ 0 & 3\%^2 & 0 \\ 0 & 0 & 1\%^2 \end{bmatrix}$$

2. The off diagonal elements of $\mathbf{\Omega}$ are typically set to zero. This is because it is often convenient to assume (but of course not necessary) that views are uncorrelated.

3. The larger the values of the diagonal elements, the more the investor's predicted expected return can vary, and the less confidence there is in your views.

$\triangle$

Estimate Market Equilibrium, π

Example 7.7. Our starting point is CAPM. From Section 5.3, we obtain that as $\delta = \frac{\mathbb{E}[r_M] - r_f}{\sigma_M^2}$ is defined to be the "market price of risk", the expected excess return on security i becomes

$$\pi = \delta \mathbf{\Sigma} \mathbf{w}_M$$

Remark 7.8. Given market capitalization weights, market price of risk and a covariance matrix

- Determining π by calculating $\pi = \delta \mathbf{\Sigma} w_M$ is called "reverse optimization"

- π is referred to as the "market implied expected returns" or "equilibrium implied expected returns"

$\triangle$

7.2   Theil's Mixed Estimation Formulation

Calculating Black-Litterman Expected Returns

We can rewrite assumption (b) and (c) as

$$
\begin{aligned}
\pi &= \mu + \nu, \quad \nu \sim \mathcal{N}(\mathbf{0}, \tau\mathbf{\Sigma}) \in \mathbb{R}^n \\
q &= \mathbf{P}\mu + \varepsilon, \quad \varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}) \in \mathbb{R}^k
\end{aligned}
\quad \Longleftrightarrow \quad
\begin{bmatrix} \pi \\ q \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{P} \end{bmatrix} \mu + \begin{bmatrix} \nu \\ \varepsilon \end{bmatrix}
$$

Stacking these two equations in the form

$$y = \mathbf{X}\mu + \varepsilon', \quad \varepsilon' \sim \mathcal{N}(0, \mathbf{V})$$

where $\mathbf{I}$ denotes the $n \times n$ identity matrix, and

$$
y = \begin{bmatrix} \pi \\ q \end{bmatrix} \in \mathbb{R}^{n+k}, \quad
\mathbf{X} = \begin{bmatrix} \mathbf{I} \\ \mathbf{P} \end{bmatrix} \in \mathbb{R}^{(n+k) \times n}, \quad
\mathbf{V} = \begin{bmatrix} \tau\mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} \end{bmatrix} \in \mathbb{R}^{(n+k) \times (n+k)}.
$$

It's easy to see that this is just a standard linear model for the expected returns $\mu = \mathbb{E}[r]$ that can be solved using generalized/weighted least squares (GLS). Calculating the GLS

estimator we obtain the Black-Litterman expected return

$$
\begin{aligned}
\hat{\boldsymbol{\mu}}_{BL} &= \left(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \boldsymbol{y} \\
&= \left(\begin{bmatrix} \mathbf{I} & \mathbf{P}^\top \end{bmatrix} \begin{bmatrix} (\tau\boldsymbol{\Sigma})^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{P} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{I} & \mathbf{P}^\top \end{bmatrix} \begin{bmatrix} (\tau\boldsymbol{\Sigma})^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\pi} \\ \boldsymbol{q} \end{bmatrix} \\
&= \left(\begin{bmatrix} \mathbf{I} & \mathbf{P}^\top \end{bmatrix} \begin{bmatrix} (\tau\boldsymbol{\Sigma})^{-1} \\ \boldsymbol{\Omega}^{-1}\mathbf{P} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{I} & \mathbf{P}^\top \end{bmatrix} \begin{bmatrix} (\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi} \\ \boldsymbol{\Omega}^{-1}\boldsymbol{q} \end{bmatrix} \\
&= \left[(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^\top \boldsymbol{\Omega}^{-1} \mathbf{P} \right]^{-1} \left[(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi} + \mathbf{P}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{q} \right]
\end{aligned}
$$

Therefore, we get

$$
\hat{\boldsymbol{\mu}}_{BL} = \left[(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^\top \boldsymbol{\Omega}^{-1} \mathbf{P} \right]^{-1} \left[(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi} + \mathbf{P}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{q} \right] \in \mathbb{R}^n
$$

This is the (posterior) Black-Litterman expected returns that "blend" the market equilibrium with the investor's views.

Remark 7.9. This way of calculating the Black-Litterman expected return is called the mixed estimation method of Theil and Goldberger (1961)[9]. △

Calculating the Covariance Matrix of Black-Litterman Expected Returns

To fit the Black-Litterman approach into the MVO framework, we still need the covariance matrix, $\boldsymbol{\Sigma}_{BL} := \mathbb{V}[\hat{\boldsymbol{\mu}}_{BL}]$, of Black-Litterman Expected Returns, $\hat{\boldsymbol{\mu}}_{BL}$.

Theorem 7.10. *(The Covariance Matrix of the Black-Litterman Returns) The covariance (i.e. uncertainty) of the Black-Litterman expected return $\hat{\boldsymbol{\mu}}_{BL}$ is*

$$
\boldsymbol{\Sigma}_{BL} := \mathbb{V}\left[\hat{\boldsymbol{\mu}}_{BL}\right] = \left[(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^\top \boldsymbol{\Omega}^{-1} \mathbf{P} \right]^{-1}
$$

Note that here "$\mathbb{V}$" stands for the covariance matrix of $\hat{\boldsymbol{\mu}}_{BL}$.

Proof. Since $\hat{\boldsymbol{\mu}}_{BL} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\boldsymbol{y}$, and $\mathbf{X}, \mathbf{V}$ can be viewed as constant matrices, we obtain by the definition of the covariance matrix

$$
\begin{aligned}
\mathbb{V}[\hat{\boldsymbol{\mu}}_{BL}] &= \mathbb{E}[(\hat{\boldsymbol{\mu}}_{BL} - \mathbb{E}[\hat{\boldsymbol{\mu}}_{BL}])(\hat{\boldsymbol{\mu}}_{BL} - \mathbb{E}[\hat{\boldsymbol{\mu}}_{BL}])^\top] \\
&= \mathbb{E}\left[\left((\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{y} - \mathbb{E}[\boldsymbol{y}])\right)\left((\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{y} - \mathbb{E}[\boldsymbol{y}])\right)^\top \right] \\
&= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbb{E}[(\boldsymbol{y} - \mathbb{E}[\boldsymbol{y}])(\boldsymbol{y} - \mathbb{E}[\boldsymbol{y}])^\top]\left((\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\right)^T \\
&= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbb{V}[\boldsymbol{y}]\left((\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\right)^T \\
\mathbb{V}[\boldsymbol{y}] = \mathbf{V} \implies &= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{V}\left((\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\right)^T \\
\mathbf{V}^\top = \mathbf{V} \implies &= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}XX^{-1}\mathbf{V}(\mathbf{X}^T)^{-1} \\
&= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} \\
&= [(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}\mathbf{P}]^{-1}
\end{aligned}
$$

Hence, we have

$$\boldsymbol{\Sigma}_{\mathbf{B}L} := \mathbb{V}[\hat{\boldsymbol{\mu}}_{BL}] = [(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}\mathbf{P}]^{-1}.$$

$\square$

Calculating the portfolio weights under the Black-Litterman framework

Under the Black-Litterman model, the Black-Litterman expected returns generated by assumption (b) and (c) is $\boldsymbol{\mu}_{BL}$. Since assumption (a) specifies that $r \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we can write r as $r = \hat{\boldsymbol{\mu}}_{\mathbf{B}L} + \boldsymbol{\xi}$ where $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Then we have $\mathbb{V}[r] = \mathbb{V}[\hat{\boldsymbol{\mu}}_{\mathbf{B}L}] + \mathbb{V}[\boldsymbol{\xi}] = \boldsymbol{\Sigma}_{BL} + \boldsymbol{\Sigma}$. That is,

$$r \sim \mathcal{N}(\hat{\boldsymbol{\mu}}_{\mathbf{B}L}, \boldsymbol{\Sigma} + \boldsymbol{\Sigma}_{\mathbf{B}L})$$

We just plug in the covariance matrix, $\mathbb{V}[r] = \boldsymbol{\Sigma}_{BL} + \boldsymbol{\Sigma}$, and the Black-Litterman Expected Returns, $\hat{\boldsymbol{\mu}}_{BL}$, to the classical MVO framework to get our optimal w^*.

$$w^* = \arg\min_{\boldsymbol{w}}\left\{\boldsymbol{w}^\top\hat{\boldsymbol{\mu}}_{BL} - \frac{\lambda}{2}\boldsymbol{w}^\top\mathbb{V}[r]\boldsymbol{w}\right\} = \lambda^{-1}(\boldsymbol{\Sigma}_{BL} + \boldsymbol{\Sigma})^{-1}\hat{\boldsymbol{\mu}}_{BL}$$

Remark 7.11. Moreover, we have

$$\mathbb{V}[r] = \boldsymbol{\Sigma} + \boldsymbol{\Sigma}_{\mathbf{B}L}$$
$$= \boldsymbol{\Sigma} + \left[(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^\top\boldsymbol{\Omega}^{-1}\mathbf{P}\right]^{-1}$$
$$\textit{Corollary } 2.11 \implies = \boldsymbol{\Sigma} + \tau\boldsymbol{\Sigma} - \tau\boldsymbol{\Sigma}\mathbf{P}^\top(\boldsymbol{\Omega} + \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top)^{-1}\mathbf{P}\tau\boldsymbol{\Sigma}$$
$$= (1+\tau)\boldsymbol{\Sigma} - \tau^2\boldsymbol{\Sigma}\mathbf{P}^\top\left[\mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top + \boldsymbol{\Omega}\right]^{-1}\mathbf{P}\boldsymbol{\Sigma}$$

It's easy to see that

- $\mathbb{V}[r] > \boldsymbol{\Sigma}$ and $\mathbb{V}[r] > \boldsymbol{\Sigma}_{\mathbf{B}L}$

- $\tau = 0$ (no estimation risk) then $\mathbb{V}[r] \equiv \boldsymbol{\Sigma}$

- $\boldsymbol{\Omega} \to \infty$ (uninformative prior) then $\boldsymbol{\Sigma}_{\mathbf{B}L} \equiv (1+\tau)\boldsymbol{\Sigma}$

- Note that if $\boldsymbol{\Omega} = \alpha \cdot \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top$ (with $\mathbf{P}$ full rank) then $\mathbb{V}[r] \equiv \left(1 + \tau - \frac{\tau}{1+\alpha}\right)\boldsymbol{\Sigma}$

$\triangle$

7.3   Bayesian Formulation

We derive it as in Section 2.4.

(a) Set parameter $\boldsymbol{\theta} = \mathbb{E}[r]$, and

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

Note that this can be viewed as $r = \boldsymbol{\theta} + \varepsilon$ where $\varepsilon \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

(b) Set prior $\pi(\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\pi}, \mathbf{C})$, where $\boldsymbol{\pi} = \delta\boldsymbol{\Sigma}\boldsymbol{w}_M$ is the market implied expected return. That is,

$$\pi(\boldsymbol{\theta}) = \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\pi})^\top \mathbf{C}^{-1}(\boldsymbol{\theta} - \boldsymbol{\pi})\right]$$

(c) The view function is given by $\mathbf{P}\boldsymbol{\theta} = \boldsymbol{q} + \boldsymbol{\nu}$ where $\boldsymbol{\nu} \sim \mathcal{N}(0, \boldsymbol{\Omega})$. Then we have $\boldsymbol{q}|\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{P}\boldsymbol{\theta}, \boldsymbol{\Omega})$, and the likelihood function is

$$f(\boldsymbol{q}|\boldsymbol{\theta}) = \exp\left[-\frac{1}{2}(\boldsymbol{q} - \mathbf{P}\boldsymbol{\theta})^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{q} - \mathbf{P}\boldsymbol{\theta})\right]$$

Solution: Then, under the Bayesian framework, we have

$$p(\boldsymbol{\theta}|\boldsymbol{q}) = f(\boldsymbol{q}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$$

$$= \exp\left[-\frac{1}{2}(\boldsymbol{q} - \mathbf{P}\boldsymbol{\theta})^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{q} - \mathbf{P}\boldsymbol{\theta})\right] \cdot \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\pi})^\top \mathbf{C}^{-1}(\boldsymbol{\theta} - \boldsymbol{\pi})\right]$$

$$= \exp\left[-\frac{1}{2}\left(\boldsymbol{\theta}^\top(\mathbf{P}\boldsymbol{\Omega}^{-1}\mathbf{P} + \mathbf{C}^{-1})\boldsymbol{\theta} - 2(\boldsymbol{q}^\top\boldsymbol{\Omega}^{-1}\mathbf{P} + \boldsymbol{\pi}^\top\mathbf{C}^{-1})\boldsymbol{\theta}\right)\right]$$

$$\implies -2\log p(\boldsymbol{\theta}|\boldsymbol{q}) = \boldsymbol{\theta}^\top(\mathbf{P}\boldsymbol{\Omega}^{-1}\mathbf{P} + \mathbf{C}^{-1})\boldsymbol{\theta} - 2(\boldsymbol{q}^\top\boldsymbol{\Omega}^{-1}\mathbf{P} + \boldsymbol{\pi}^\top\mathbf{C}^{-1})\boldsymbol{\theta}$$

Since the likelihood $f(\boldsymbol{q}|\boldsymbol{\theta})$ is normal and the prior $\pi(\boldsymbol{\theta})$ is also normal, which is a conjugate prior for that likelihood, it implies that the posterior is of the same family, i.e., $\boldsymbol{\theta}|\boldsymbol{q}$ is also normally distributed.

Set $\mathbf{M} = (\mathbf{P}\boldsymbol{\Omega}^{-1}\mathbf{P} + \mathbf{C}^{-1})$ and $\boldsymbol{u}^\top = (\boldsymbol{q}^\top\boldsymbol{\Omega}^{-1}\mathbf{P} + \boldsymbol{\pi}^\top\mathbf{C}^{-1})$ in Theorem 2.14, we have

$$\mathbf{V}[\boldsymbol{\theta}|\boldsymbol{q}] = (\mathbf{P}\boldsymbol{\Omega}^{-1}\mathbf{P} + \mathbf{C}^{-1})^{-1}, \quad \mathbb{E}[\boldsymbol{\theta}|\boldsymbol{q}] = (\mathbf{P}\boldsymbol{\Omega}^{-1}\mathbf{P} + \mathbf{C}^{-1})^{-1}(\mathbf{P}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{q} + \mathbf{C}^{-1}\boldsymbol{\pi})$$

Therefore, we obtain

$$\begin{cases} \mathbb{E}[\boldsymbol{r}|\boldsymbol{q}] = \mathbb{E}[\boldsymbol{\theta}|\boldsymbol{q}] = (\mathbf{P}\boldsymbol{\Omega}^{-1}\mathbf{P} + \mathbf{C}^{-1})^{-1}(\mathbf{P}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{q} + \mathbf{C}^{-1}\boldsymbol{\pi}), \\ \mathbf{V}[\boldsymbol{r}|\boldsymbol{q}] = \mathbf{V}[\boldsymbol{\theta}|\boldsymbol{q}] + \mathbf{V}[\varepsilon|\boldsymbol{q}] = (\mathbf{P}\boldsymbol{\Omega}^{-1}\mathbf{P} + \mathbf{C}^{-1})^{-1} + \boldsymbol{\Sigma} \end{cases}$$

We just plug in the covariance matrix, $\mathbf{V}[\varepsilon|\boldsymbol{q}]$, and the Black-Litterman Expected Returns, $\mathbb{E}[\boldsymbol{r}|\boldsymbol{q}]$, to the classical MVO framework to get our optimal $\boldsymbol{w}^*$.

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}}\left\{\boldsymbol{w}^\top\mathbb{E}[\boldsymbol{r}|\boldsymbol{q}] - \frac{\lambda}{2}\boldsymbol{w}^\top\mathbf{V}[\varepsilon|\boldsymbol{q}]\boldsymbol{w}\right\}$$

From Problem 4.33, we get that the optimal portfolio is then given by

$$\boldsymbol{w}^* = \lambda^{-1}\left[(\mathbf{P}\boldsymbol{\Omega}^{-1}\mathbf{P} + \mathbf{C}^{-1})^{-1} + \boldsymbol{\Sigma}\right]^{-1}(\mathbf{P}\boldsymbol{\Omega}^{-1}\mathbf{P} + \mathbf{C}^{-1})^{-1}(\mathbf{P}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{q} + \mathbf{C}^{-1}\boldsymbol{\pi})$$

Specifically, choose $\mathbf{C} = \tau\boldsymbol{\Sigma}$, we recover the classical BL model.

7.3.1 Roadmap of the Black-Litterman Framework

Now, we can present a clear and concise illustration of the Black-Litterman framework in the image provided below.
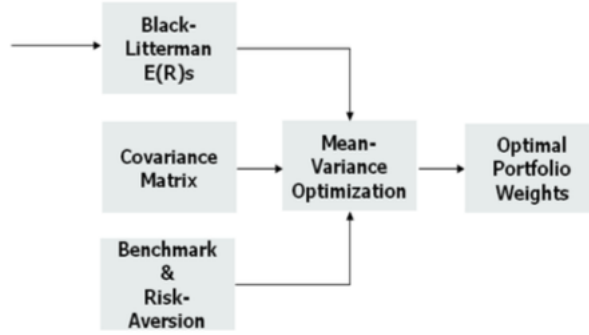

Figure 12: Roadmap of the Black-Litterman model

7.4 Some Properties of the Black-Litterman Model

In this section, we examine various properties of the Black-Litterman model to gain deeper insights and uncover the implications suggested by the model, ultimately leading to a better understanding of the Black-Litterman framework.

Property 7.12. *(**Absence of Views**) If the investor has no views (i.e. $q = 0, k = 0$) or the confidence in the views is zero (i.e. $\Omega \rightarrow \infty$), then the Black-Litterman returns are equal to the equilibrium returns, i.e. $\hat{\mu}_{\mathbf{BL}} = \pi$.*

Proof. This is relatively easy to prove.

(a) If $q = 0, k = 0$, then we have

$$\hat{\mu}_{BL} = \left[(\tau\Sigma)^{-1} \right]^{-1} (\tau\Sigma)^{-1}\pi = \pi$$

(b) If $\Omega \rightarrow \infty$, we get

$$\hat{\mu}_{BL} = \left[(\tau\Sigma)^{-1} + \mathbf{P}^{\top}\Omega^{-1}\mathbf{P} \right]^{-1} \left[(\tau\Sigma)^{-1}\pi + \mathbf{P}^{\top}\Omega^{-1}q \right] = \left[(\tau\Sigma)^{-1} \right]^{-1} (\tau\Sigma)^{-1}\pi = \pi$$

$\square$

Remark 7.13. (Interpretation)

1. Consequently, with no views ($q = 0$) or views with of zero confidence ($\Omega \rightarrow \infty$), the investor will end up holding the market portfolio. In other words, the optimal portfolio in the absence of views is the defined market.

$\triangle$

Property 7.14. *(**"Confidence" Weighted Expected Return**) We can rewrite the Black-Litterman expected returns in the form*

$$\hat{\boldsymbol{\mu}}_{BL} = \left[(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^\top\boldsymbol{\Omega}^{-1}\mathbf{P}\right]^{-1}\left[(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi} + \mathbf{P}^\top\boldsymbol{\Omega}^{-1}\mathbf{P}\hat{\boldsymbol{\mu}}_{investor}\right]$$
$$= w_{\boldsymbol{\pi}}\boldsymbol{\pi} + w_{\boldsymbol{q}}\hat{\boldsymbol{\mu}}_{investor}$$

where the two weighting matrices are given by

$$w_{\boldsymbol{\pi}} = \left[(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^\top\boldsymbol{\Omega}^{-1}\mathbf{P}\right]^{-1}(\tau\boldsymbol{\Sigma})^{-1}$$
$$w_{\boldsymbol{q}} = \left[(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^\top\boldsymbol{\Omega}^{-1}\mathbf{P}\right]^{-1}\mathbf{P}^\top\boldsymbol{\Omega}^{-1}\mathbf{P} \qquad \Longrightarrow \qquad w_{\boldsymbol{\pi}} + w_{\boldsymbol{q}} = \mathbf{I} \in \mathbb{R}^{n\times n}$$

Proof. By using $\boldsymbol{q} = \mathbf{P}\boldsymbol{\mu} + \boldsymbol{\varepsilon}$, we have that the investor's views alone imply the estimate of expected returns

$$\hat{\boldsymbol{\mu}}_{\text{investor}} = \left(\mathbf{P}^\top\mathbf{P}\right)^{-1}\mathbf{P}^\top\boldsymbol{q}.$$

Since $\mathbf{P}\left(\mathbf{P}^\top\mathbf{P}\right)^{-1}\mathbf{P}^\top = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix, we must have $\mathbf{P}\hat{\boldsymbol{\mu}}_{\text{investor}} = \boldsymbol{q}$. Therefore, we have

$$\hat{\boldsymbol{\mu}}_{BL} = \left[(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^\top\boldsymbol{\Omega}^{-1}\mathbf{P}\right]^{-1}\left[(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi} + \mathbf{P}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{q}\right]$$
$$= \left[(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^\top\boldsymbol{\Omega}^{-1}\mathbf{P}\right]^{-1}\left[(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi} + \mathbf{P}^\top\boldsymbol{\Omega}^{-1}\mathbf{P}\hat{\boldsymbol{\mu}}_{\text{investor}}\right]$$

$\square$

Remark 7.15. (Interpretation:)

1. Notice, $\hat{\boldsymbol{\mu}}_{\text{investor}}$ and $\hat{\boldsymbol{\mu}}_{BL}$ are different.

2. We see that the Black-Litterman expected return, $\hat{\boldsymbol{\mu}}_{BL} = w_{\boldsymbol{\pi}}\boldsymbol{\pi} + w_{\boldsymbol{q}}\hat{\boldsymbol{\mu}}_{\text{investor}}$, is a "confidence weighted" linear combination (or "weighted average")of market equilibrium $\boldsymbol{\pi}$ and the expected return $\hat{\boldsymbol{\mu}}_{\text{investor}}$ implied by the investor's views.

3. In particular, $(\tau\boldsymbol{\Sigma})^{-1}$ and $\mathbf{P}^\top\boldsymbol{\Omega}^{-1}\mathbf{P}$ represent the **confidence** we have in our estimates of the market equilibrium and the views, respectively.

   Therefore, if we have low confidence in the views, the resulting expected returns will be close to the ones implied by market equilibrium — $\boldsymbol{\pi}$. Conversely, with higher confidence in the views, the resulting expected returns will deviate from the market and be close to investor's views — $\boldsymbol{q}$.

$\triangle$

Property 7.16. *(**Deviation from Market Equilibrium**) The Black-Litterman expected return $\hat{\boldsymbol{\mu}}_{BL}$ can be rewritten as*

$$\hat{\boldsymbol{\mu}}_{BL} = \boldsymbol{\pi} + \tau\boldsymbol{\Sigma}\mathbf{P}^\top\left[\mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top + \boldsymbol{\Omega}\right]^{-1}\left[\boldsymbol{q} - \mathbf{P}\boldsymbol{\pi}\right]$$

Proof. Set $\mathbf{A} = \boldsymbol{\Omega}$ and $\mathbf{B} = \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^T$ in Lemma 2.3, we get

$$(\boldsymbol{\Omega} + \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top)^{-1} = \boldsymbol{\Omega}^{-1}\left(\mathbf{I} + \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top\boldsymbol{\Omega}^{-1}\right)^{-1}$$

$$\implies (\boldsymbol{\Omega} + \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top)^{-1}\left(\mathbf{I} + \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top\boldsymbol{\Omega}^{-1}\right) = \boldsymbol{\Omega}^{-1}$$

$$\implies (\boldsymbol{\Omega} + \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top)^{-1} = \boldsymbol{\Omega}^{-1} - (\boldsymbol{\Omega} + \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top)^{-1}\mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top\boldsymbol{\Omega}^{-1} \quad (24)$$

Plug the result from Corollary 2.11 into the formula for $\hat{\boldsymbol{\mu}}_{BL}$, we have

$$\begin{aligned}
\hat{\boldsymbol{\mu}}_{BL} &= [(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}\mathbf{P}]^{-1}[(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}q] \\
&= \left(\tau\boldsymbol{\Sigma} - \tau\boldsymbol{\Sigma}\mathbf{P}^\top(\boldsymbol{\Omega} + \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top)^{-1}\mathbf{P}\tau\boldsymbol{\Sigma}\right)[(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}q] \\
&= \boldsymbol{\pi} - \tau\boldsymbol{\Sigma}\mathbf{P}^T(\boldsymbol{\Omega} + \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top)^{-1}\mathbf{P}\boldsymbol{\pi} + \tau\boldsymbol{\Sigma}\mathbf{P}^T\boldsymbol{\Omega}^{-1}q \\
&\quad - \tau\boldsymbol{\Sigma}\mathbf{P}^\top(\boldsymbol{\Omega} + \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top)^{-1}\mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^T\boldsymbol{\Omega}^{-1}q \\
&= \boldsymbol{\pi} - \tau\boldsymbol{\Sigma}\mathbf{P}^T(\boldsymbol{\Omega} + \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top)^{-1}\mathbf{P}\boldsymbol{\pi} \\
&\quad + \tau\boldsymbol{\Sigma}\mathbf{P}^T\left(\boldsymbol{\Omega}^{-1} - (\boldsymbol{\Omega} + \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top)^{-1}\mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^T\boldsymbol{\Omega}^{-1}\right)q \\
\text{Equation (24)} \implies &= \boldsymbol{\pi} - \tau\boldsymbol{\Sigma}\mathbf{P}^T(\boldsymbol{\Omega} + \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top)^{-1}\mathbf{P}\boldsymbol{\pi} + \tau\boldsymbol{\Sigma}\mathbf{P}^T(\boldsymbol{\Omega} + \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top)^{-1}q \\
&= \boldsymbol{\pi} + \tau\boldsymbol{\Sigma}\mathbf{P}^T(\boldsymbol{\Omega} + \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top)^{-1}[q - \mathbf{P}\boldsymbol{\pi}]
\end{aligned}$$

Therefore, we have

$$\hat{\boldsymbol{\mu}}_{BL} = \boldsymbol{\pi} + \tau\boldsymbol{\Sigma}\mathbf{P}^\top\left[\mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top + \boldsymbol{\Omega}\right]^{-1}[q - \mathbf{P}\boldsymbol{\pi}]$$

$\square$

Remark 7.17. (Interpretation:) The Black-Litterman expected return $\hat{\boldsymbol{\mu}}_{BL}$ "tilts" away from the market equilibrium is given by a vector proportional to $\boldsymbol{\Sigma}\mathbf{P}^T[\mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^T + \boldsymbol{\Omega}]^{-1}[q - \mathbf{P}\boldsymbol{\pi}]$. $\triangle$

Property 7.18. *(Optimization Formulation) The Black-Litterman expected returns can be shown to be the solution to the optimization problem*

$$\hat{\boldsymbol{\mu}}_{BL} = \underset{\boldsymbol{\mu}}{\arg\min}\left\{(\boldsymbol{\pi} - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{\pi} - \boldsymbol{\mu}) + \tau(q - \mathbf{P}\boldsymbol{\mu})^\top\boldsymbol{\Omega}^{-1}(q - \mathbf{P}\boldsymbol{\mu})\right\}$$

Remark 7.19. (Interpretation:) From this formulation we see that

1. $\hat{\boldsymbol{\mu}}_{BL}$ is chosen such that it is simultaneously as close to $\boldsymbol{\pi}$, and $\mathbf{P}\boldsymbol{\mu}$ is as close to q as possible. The distances are determined by $\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Omega}^{-1}$.

2. The relative importance of the equilibrium versus the views is determined by τ. For example, for τ large the weight of the views is increased, whereas for τ small the weight of the equilibrium is higher. (Can be viewed as a penalty term.)

3. Furthermore, we also see that τ is a "redundant" parameter as it can be absorbed into $\boldsymbol{\Omega}$.

$\triangle$

7.5   Pros and Cons of the BL Model

Still, we will use the data presented in Section 3 to better illustrate our points.

We use the market portfolio as the benchmark, where we can use the "Equilibrium Market return (%)" and the "Equilibrium Portfolio Weight (%)" to construct it.

For the Black-Litterman framework, we then suppose we have the view that:

- The German market will outperform the French market by 20%.

- The German market will outperform the UK market by 20%.

Therefore, by our Black-Litterman construction, we have the pick matrix, view matrix, and residual as:

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 \end{bmatrix}, \boldsymbol{q} = \begin{bmatrix} 20\% \\ 20\% \end{bmatrix}, \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}), \boldsymbol{\Omega} = \begin{bmatrix} 0.04 & 0 \\ 0 & 0.04 \end{bmatrix}$$

Plug these into the BL framework and let's compare the market portfolio weights and the portfolio weights suggested by the Black-Litterman model incorporated with our views.
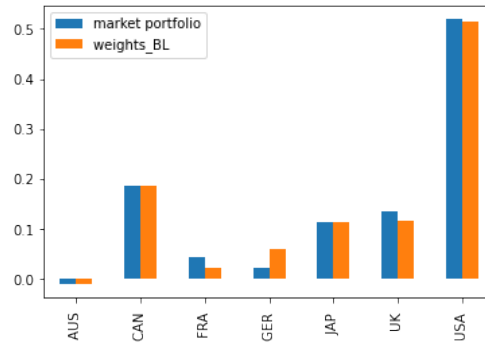
Figure 13: Market portfolio weigths vs. optimal BL weights

From Figure 13 presented above, it is evident that the portfolio weights suggested by the BL model exhibit improved interpretability. As we are more optimistic about Germany, its weight is higher, while the weights of the UK and France are lower. This allocation aligns with investor expectations and is more easily accepted compared to the MVO framework.

The weights of other countries remain equal to those in the market equilibrium state, unaffected by our subjective judgments, which echo back the idea behind the BL framework where the investor maintains the market portfolio if they don't have views. Such mechanics increase the lower bound of the return of the portfolio yields by the BL model and add to the robustness of the weights provided by the BL framework.

Pros:

1. Better interpretability comparing to the classical MVO framework.

2. More stable/robust comparing to the classical MVO.

3. It's relatively simple. The investor is now no longer obliged to specify views on all assets, but can specify views only if she holds one.

Cons:

1. The BL model assumes that security returns are normally distributed, which is not realistic empirically speaking.

2. The assumption that investment views are linear portfolios of the securities excludes the use of nonlinear views that are pervasive in options strategies, such as tail-risk hedging overlays.

3. The prior distribution in the classical BL model is provided by the CAPM market equilibrium portfolio, where a number of factors, in addition to the market factor, are important for the derivation of prior returns.

4. The confidence of views, represented by τ, and the risk-aversion coefficient, represented by δ, are difficult to calibrate. We can only rely on rules of thumb for their determination.

7.6   Practical Considerations about the Black-Litterman Model

7.6.1   What Should the Magnitude of τ Be?

Recall that

$$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\pi}, \tau\boldsymbol{\Sigma}) \quad \text{(market equilibrium)}$$

The sample estimator

$$\hat{\boldsymbol{\mu}} = \frac{1}{T} \sum_{t=1}^{T} r_i \sim \mathcal{N}\left(\boldsymbol{\pi}, \frac{1}{T}\boldsymbol{\Sigma}\right)$$

Therefore it is sensible that $\tau \propto \frac{1}{T}$.

7.6.2   Calibrating the Confidence in the Views, $\boldsymbol{\Omega}$

$$\mathbf{P}\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{q}, \boldsymbol{\Omega}) \quad \text{(investor's views)}$$

In many cases we may not have a direct estimate of the confidence (variance) of a view, so we have to apply other tools to find it out. We have several ways to do this:

1. Using the prior $\boldsymbol{\Omega} = \mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top$ or $\boldsymbol{\Omega} = \text{diag}\left(\mathbf{P}\tau\boldsymbol{\Sigma}\mathbf{P}^\top\right)$.

2. Using confidence intervals like we did in the example above.

   Remark 7.20. Let us say that we believe there is a 95% chance that the first view in Section 7.1 will happen. If we assume normality, we can interpret this as a 95% confidence interval for the future return to be in the interval $[3\% - 3.9\%, 3\% +$

3.9%]. From this confidence interval we calculate that the implied standard deviation (σ_1) is equal to about 2%. Therefore, we would set the Black-Litterman confidence equal to $(2\%)^2 = 0.04\%$. $\triangle$

3. From backtesting the trading strategies (*e.g.* long-short strategy) corresponding to the views.

   Remark 7.21. We could estimate its historical variance through simulation with historical data. Of course, we cannot completely judge the performance of a strategy going forward from our backtests. Nevertheless, the back-test methodology allows us to obtain an estimate of the Black-Litterman view and confidence for a particular view/strategy. $\triangle$

4. Factor models (cross-sectional or time-series).

   Remark 7.22. APT models are especially useful as they provide a covariance matrix estimator that, unlike the sample covariance matrix, leads to stable, well-diversified optimal portfolios. This is provided when you have great risk factors, solid estimation techniques, and a big, well-constructed portfolio. $\triangle$

7.7   Connecting the Black-Litterman model with APT

Intuition: Our goal is to determine how our view, denoted as q, influences the distribution of r, specifically $\mathbb{E}[r]$ and $\mathbb{V}[r]$. To achieve this, we introduce a latent parameter, θ, as an intermediary. This parameter could represent $\mathbb{E}[r]$ or the expected value of other factors related to r. We first consider the prior of θ, then apply Bayesian techniques to update θ and obtain the posterior $\theta|q$. Finally, we use this posterior of θ to update the distribution of r, employing $r|\theta \cdot \theta|q$.

This is proposed by Petter Klom and Gordon Ritter(2017).

7.7.1   Formulation

Motivation: Instead of analyzing the distribution of parameter $\theta = \mu$, the expected return of r directly as in the classical Black-Litterman Framework, i.e.

$$r \xleftarrow{\text{Bayesian}} q,$$

we investigate the distribution of expected factor risk premia $\theta = \mathbb{E}[f]$, then use the APT model to bridge the gap between r and f, i.e.

$$r \underset{r|q \quad \Longleftarrow \quad r|\theta\cdot\theta|q}{\xleftarrow{\text{Bayesian}}} \theta|q \underset{\theta|q \quad \Longleftarrow \quad q|\theta\cdot\pi(\theta)}{\xleftarrow{\text{Bayesian}}} q.$$

Definition 7.23. *(Essential Ingredients)*

(a) A parametric statistical model for asset returns $p(r|\theta)$ with finite-dimensional parameter vector θ,

(b) A prior $\pi(\theta)$ on the parameter space,

(c) A likelihood function $f(q|\theta)$ where θ is any parameter vector appearing in a parametric statistical model for asset returns, and q is a vector supplied by portfolio managers or economists.

(d) Explicitly, the posterior predictive density of r is given by

$$p(r|q) = \int p(r|\theta)p(\theta|q)d\theta \quad \text{where}$$

$$p(\theta|q) = \frac{f(q|\theta)\pi(\theta)}{\int f(q|\theta)\pi(\theta)d\theta}$$

Remark 7.24.

1. Items (a) and (b) simply state that we have a Bayesian statistical model, as defined in Section 2.4, for asset returns.

2. We are free to choose θ as any vector of parameters appearing in a parametric statistical model for asset returns.

$\triangle$

Step 0: Introduce the APT model.

From Section 6.2, we know that the APT model states

$$r = \mathbf{X}f + \varepsilon, \quad \mathbb{E}[\varepsilon] = 0, \quad \mathbb{V}[\varepsilon] = \mathbf{D}$$

where

- $r \in \mathbb{R}^n$ is an n-dimensional random vector containing the crosssection of **returns in excess of the risk-free rate** over some time interval $[t, t+1]$,

- $\mathbf{X} \in \mathbb{R}^{n \times k}$ is a (non-random) $n \times k$ matrix for factor loadings that is known before time t,

- $f \in \mathbb{R}^k$ denotes a k-dimensional random vector process for latent factors which cannot be observed directly,

- $k \ll n$,

- $\varepsilon \in \mathbb{R}^n$ is assumed to follow a mean-zero distribution with diagonal variance-covariance matrix given by $\mathbf{D} = diag\left(\sigma_1^2, \ldots, \sigma_n^2\right)$ with all $\sigma_i^2 > 0$.

Also from Proposition 6.4, we know that if the f-process is assumed to have finite first and second moments as

$$\mathbb{E}[f] = \mu_f, \text{ and } \mathbb{V}[f] = \mathbf{F}$$

then

$$\mathbb{E}[r] = \mathbf{X}\mu_f, \text{ and } \mathbb{V}[r] = \mathbf{D} + \mathbf{X}\mathbf{F}\mathbf{X}^\top \tag{25}$$

where $\mathbf{X}^\top$ denotes the transpose of $\mathbf{X}$.

Step 1: Specify parameter $\boldsymbol{\theta}$.

We are free to choose $\boldsymbol{\theta}$ as any vector of parameters appearing in a parametric statistical model for asset returns. Therefore, we use the APT model and choose

$$\boldsymbol{\theta} = \boldsymbol{\mu}_f, \text{ i.e. } \boldsymbol{f} \sim \mathcal{N}(\boldsymbol{\theta}, \mathbf{F})$$

which represents the k parameters describing the factor risk premia.

For simplicity we treat $\mathbf{F}$ as a constant matrix, just as the original Black-Litterman model treats $\boldsymbol{\Sigma}$ as a constant matrix.

Step 2: Choose the prior $\pi(\boldsymbol{\theta})$ for parameter $\boldsymbol{\theta}$.

- **Option 1:** Data-Driven Prior.

  In particular, we pick the historical mean and variance of the OLS estimates $\hat{\boldsymbol{f}}_t = \left(\mathbf{X}_t^\top \mathbf{X}_t\right)^{-1} \mathbf{X}_t^\top \boldsymbol{r}_{t+1}$ to be taken as the prior mean and prior variance.

- **Option 2:** Benchmark-Optimal Prior.

  We choose prior such that $\pi(\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\xi}, \boldsymbol{V})$ with $\boldsymbol{\xi} \in \mathbb{R}^k$ and symmetric positive definite matrix $\boldsymbol{V} \in \mathbb{R}^{k \times k}$. At this prior state, we have a prior distribution of $\boldsymbol{r}$ given by $p(\boldsymbol{r}) = \int p(\boldsymbol{r}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ where $p(\boldsymbol{r}|\boldsymbol{\theta})$ is given by the APT model. Under this distribution $p(\boldsymbol{r})$, we have the corresponding $\mathbb{E}[\boldsymbol{r}]$ and $\mathbf{V}[\boldsymbol{r}]$, which yields a optimal portfolio weights under a specified utility function given by the MVO framework, and we call this portfolio the benchmark portfolio $\boldsymbol{w}_B$, or the priori optimal portfolio.

Step 3: Calculate the optimal portfolio $\boldsymbol{w}^*$, along with its mean and variance, at the prior state.

Now, we treat $\pi(\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\xi}, \boldsymbol{V})$ as if the expected return $\boldsymbol{\xi}$ and covariance $\boldsymbol{V}$ are known. Let's calculate the priori optimal portfolio/benchmark portfolio $\boldsymbol{w}_B$.

Solution: The first step is to compute the a priori density of $\boldsymbol{r}$, $p(\boldsymbol{r}) = \int p(\boldsymbol{r}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$. Since $\pi(\boldsymbol{\theta})$ and $p(\boldsymbol{r}|\boldsymbol{\theta})$ are both Gaussian, this is another completion of squares.

For simplicity, we define the asset-level covariance in the APT model as

$$\boldsymbol{\Sigma} := \mathbb{V}[\boldsymbol{r}] = \mathbf{D} + \mathbf{X}\mathbf{F}\mathbf{X}^\top$$

Then, we have $\boldsymbol{r} \mid \boldsymbol{\theta} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\theta}, \boldsymbol{\Sigma})$, and

$$p(\boldsymbol{r} \mid \boldsymbol{\theta}) = \exp\left[-\frac{1}{2}(\boldsymbol{r} - \mathbf{X}\boldsymbol{\theta})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{r} - \mathbf{X}\boldsymbol{\theta})\right], \pi(\boldsymbol{\theta}) = \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\xi})^\top \mathbf{V}^{-1}(\boldsymbol{\theta} - \boldsymbol{\xi})\right]$$

Straightforward calculations then show:

$$p(r \mid \theta)\pi(\theta) = \exp\left[-\frac{1}{2}(r - \mathbf{X}\theta)^\top \Sigma^{-1}(r - \mathbf{X}\theta)\right] \cdot \exp\left[-\frac{1}{2}(\theta - \xi)^\top \mathbf{V}^{-1}(\theta - \xi)\right]$$

$$= \exp\left[-\frac{1}{2}\left((r - \mathbf{X}\theta)^\top \Sigma^{-1}(r - \mathbf{X}\theta) + (\theta - \xi)^\top \mathbf{V}^{-1}(\theta - \xi)\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\theta^\top \mathbf{H}\theta - 2\eta^\top \theta + z\right)\right]$$

where the auxiliary variables

$$\mathbf{H} = V^{-1} + \mathbf{X}^\top \Sigma^{-1}\mathbf{X}, \quad \eta = \left(\xi^\top V^{-1} + r^\top \Sigma^{-1}\mathbf{X}\right)^\top, \quad z = r^\top \Sigma^{-1}r + \xi^\top V^{-1}\xi$$

Completing the square again,

$$\theta^\top \mathbf{H}\theta - 2\eta^\top \theta + z = (\theta - \nu)^\top \mathbf{H}(\theta - \nu) - \nu^\top \mathbf{H}\nu + z, \quad \nu = \mathbf{H}^{-1}\eta$$

Hence,

$$\int p(r|\theta)\pi(\theta)d\theta = \int \exp\left[-\frac{1}{2}\left(\theta^\top \mathbf{H}\theta - 2\eta^\top \theta + z\right)\right]d\theta$$

$$= \int \exp\left[-\frac{1}{2}\left((\theta - \nu)^\top \mathbf{H}(\theta - \nu) - \nu^\top \mathbf{H}\nu + z\right)\right]d\theta$$

$$= \int \exp\left[-\frac{1}{2}(\theta - \nu)^\top \mathbf{H}(\theta - \nu)\right]d\theta \cdot \exp\left[-\frac{1}{2}\left(z - \nu^\top \mathbf{H}\nu\right)\right]$$

$$\text{Gaussian} \implies = \sqrt{(2\pi)^k |\mathbf{H}^{-1}|}\exp\left[-\frac{1}{2}\left(z - \nu^\top \mathbf{H}\nu\right)\right]$$

$$\nu = \mathbf{H}^{-1}\eta \implies = \sqrt{\frac{(2\pi)^k}{\det \mathbf{H}}}\exp\left[-\frac{1}{2}\left(z - \eta^\top \mathbf{H}^{-1}\eta\right)\right]$$

$$= \sqrt{\frac{(2\pi)^k}{\det \mathbf{H}}}\exp\left[-\frac{1}{2}(r^\top \Sigma^{-1}r + \xi^\top V^{-1}\xi - (\xi^\top V^{-1}\right.$$

$$\left. + r^\top \Sigma^{-1}\mathbf{X})\mathbf{H}^{-1}\left(V^{-1}\xi + \mathbf{X}^\top \Sigma^{-1}r\right))\right]$$

$$= \sqrt{\frac{(2\pi)^k}{\det \mathbf{H}}}\exp\left[-\frac{1}{2}\left(r^\top \left(\Sigma^{-1} - \Sigma^{-1}\mathbf{X}\mathbf{H}^{-1}\mathbf{X}^\top \Sigma^{-1}\right)r\right.\right.$$

$$\left.\left. - 2\xi^\top (\mathbf{H}V)^{-1}\mathbf{X}^\top \Sigma^{-1}r + \xi^\top (V^{-1} - (V\mathbf{H}V)^{-1})\xi)\right]$$

Therefore, we have

$$\implies -2\log p(r) = r^\top \left(\Sigma^{-1} - \Sigma^{-1}\mathbf{X}\mathbf{H}^{-1}\mathbf{X}^\top \Sigma^{-1}\right)r - 2\xi^\top (\mathbf{H}V)^{-1}\mathbf{X}^\top \Sigma^{-1}r$$

$$+ (\text{ terms without } r)$$

Note that $\int p(r \mid \theta)\pi(\theta)\mathrm{d}\theta$ is a Gaussian probability distribution for the random vector r. Since matrices $\mathbf{V}$ and $\mathbf{D}$ are symmetric, matrices Σ and $\mathbf{H}$ are symmetric.

Hence, we can set $\mathbf{M} = \left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{H}^{-1}\mathbf{X}^{\top}\boldsymbol{\Sigma}^{-1}\right)$ and $\boldsymbol{u}^{\top} = \boldsymbol{\xi}^{\top}(\mathbf{H}\boldsymbol{V})^{-1}\mathbf{X}^{\top}\boldsymbol{\Sigma}^{-1}$ in Theorem 2.14. Then, we have

$$\mathbb{V}_{\pi}[r] = \mathbf{M}^{-1} = \left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{H}^{-1}\mathbf{X}^{\top}\boldsymbol{\Sigma}^{-1}\right)^{-1}$$

and the mean is

$$\mathbb{E}_{\pi}[r] = \mathbf{M}^{-1}\boldsymbol{u} = \mathbb{V}_{\pi}[r]\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{H}^{-1}\boldsymbol{V}^{-1}\boldsymbol{\xi}$$
$$= \left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{H}^{-1}\mathbf{X}^{\top}\boldsymbol{\Sigma}^{-1}\right)^{-1}\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{H}^{-1}\boldsymbol{V}^{-1}\boldsymbol{\xi}$$

The a priori optimal portfolio under CARA utility is then

$$\boldsymbol{w}_B := \boldsymbol{w}^* = \left(\lambda \mathbb{V}_{\pi}[r]\right)^{-1}\mathbb{E}_{\pi}[r] = \lambda^{-1}\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{H}^{-1}\boldsymbol{V}^{-1}\boldsymbol{\xi} \qquad (26)$$

Remark 7.25.

1. At this prior state, we don't need q. The view vector q is for us to update the distributions of θ and r; that is, it's for the posterior.

<div align="right">△</div>

Step 4: Specify the view equation.

A typical situation is when a quantitative portfolio manager has opinions on each factor risk premium, like the value premium or momentum premium. These opinions usually don't depend on their views on other factors. It's unusual, or almost impossible, for managers to have opinions on things like the total or difference of the value and momentum premiums, or more broadly on "portfolios of factors." So, we interpret the view equation as follows:

$$\boldsymbol{\theta} = \boldsymbol{q} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}), \quad \boldsymbol{\Omega} = \mathrm{diag}\left(\omega_1^2, \ldots, \omega_k^2\right)$$

then, $q|\theta \sim \mathcal{N}(q, \boldsymbol{\Omega})$ and the likelihood function is therefore

$$f(\boldsymbol{q}|\boldsymbol{\theta}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{q})^{\top}\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta} - \boldsymbol{q})\right] = \prod_{i=1}^{k}\exp\left[-\frac{1}{2\omega_i^2}\left(\theta_i - q_i\right)^2\right]$$

Step 5: Calculate the posterior distribution for $r|q$.

1. Calculate the posterior distribution for $\theta|q$, after the views are taken into account, which is given by

$$p(\boldsymbol{\theta} \mid \boldsymbol{q}) = \frac{f(\boldsymbol{q} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\boldsymbol{q} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \propto f(\boldsymbol{q} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})$$

Solution: Since $\pi(\boldsymbol{\theta})$ is normal and the likelihood $f(\boldsymbol{q}|\boldsymbol{\theta})$ is also normal, the prior is a conjugate prior for the normal likelihood. Then, the posterior distribution $p(\boldsymbol{\theta} \mid \boldsymbol{q})$ is also normal.

Using the relation above, we have

$$p(\boldsymbol{\theta} \mid \boldsymbol{q}) \propto f(\boldsymbol{q} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})$$
$$= \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{q})^{\top}\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta} - \boldsymbol{q})\right]\exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\xi})^{\top}\mathbf{V}^{-1}(\boldsymbol{\theta} - \boldsymbol{\xi})\right]$$

Therefore, we have

$$\implies -2\log p(\boldsymbol{\theta} \mid \boldsymbol{q}) = (\boldsymbol{\theta} - \boldsymbol{q})^{\top}\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta} - \boldsymbol{q}) + (\boldsymbol{\theta} - \boldsymbol{\xi})^{\top}\mathbf{V}^{-1}(\boldsymbol{\theta} - \boldsymbol{\xi})$$
$$= \boldsymbol{\theta}^{\top}(\boldsymbol{\Omega}^{-1} + \mathbf{V}^{-1})\boldsymbol{\theta} - 2(\boldsymbol{q}^{\top}\boldsymbol{\Omega}^{-1} + \boldsymbol{\xi}^{\top}\mathbf{V}^{-1})\boldsymbol{\theta}$$
$$+ (\text{ terms without } \boldsymbol{\theta})$$

We set $\mathbf{M} = (\boldsymbol{\Omega}^{-1} + \mathbf{V}^{-1})$ and $\boldsymbol{u}^{\top} = (\boldsymbol{q}^{\top}\boldsymbol{\Omega}^{-1} + \boldsymbol{\xi}^{\top}\mathbf{V}^{-1})$ in Theorem 2.14. Then, we have

$$\mathbb{V}[\boldsymbol{\theta} \mid \boldsymbol{q}] = \mathbf{M}^{-1} = (\boldsymbol{\Omega}^{-1} + \mathbf{V}^{-1})^{-1}$$

and the mean is

$$\mathbb{E}[\boldsymbol{\theta} \mid \boldsymbol{q}] = \mathbf{M}^{-1}\boldsymbol{u} = \mathbb{V}[\boldsymbol{\theta} \mid \boldsymbol{q}]\left(\boldsymbol{V}^{-1}\boldsymbol{\xi} + \boldsymbol{\Omega}^{-1}\boldsymbol{q}\right)$$
$$= \left(\boldsymbol{V}^{-1} + \boldsymbol{\Omega}^{-1}\right)^{-1}\left(\boldsymbol{V}^{-1}\boldsymbol{\xi} + \boldsymbol{\Omega}^{-1}\boldsymbol{q}\right)$$

Therefore, the posterior must satisfy

$$\boldsymbol{\theta} \mid \boldsymbol{q} \sim \mathcal{N}(\widetilde{\boldsymbol{\xi}}, \widetilde{\boldsymbol{V}})$$

where $\widetilde{\boldsymbol{V}} = \left(\boldsymbol{V}^{-1} + \boldsymbol{\Omega}^{-1}\right)^{-1}, \quad \widetilde{\boldsymbol{\xi}} = \left(\boldsymbol{V}^{-1} + \boldsymbol{\Omega}^{-1}\right)^{-1}\left(\boldsymbol{V}^{-1}\boldsymbol{\xi} + \boldsymbol{\Omega}^{-1}\boldsymbol{q}\right)$

2. Calculate the a posteriori distribution of asset returns (also called the posterior predictive density), given by

$$p(\boldsymbol{r} \mid \boldsymbol{q}) = \int p(\boldsymbol{r} \mid \boldsymbol{\theta})p(\boldsymbol{\theta} \mid \boldsymbol{q})d\boldsymbol{\theta}$$

Solution: For this one, we can view $p(\boldsymbol{\theta} \mid \boldsymbol{q})$ as the prior density for $\boldsymbol{\theta}$, then the situation here is the same as in **Step 3**. Hence, use the result in **Step 3** and make the substitution $\boldsymbol{\xi} \to \widetilde{\boldsymbol{\xi}}$ and $\boldsymbol{V} \to \widetilde{\boldsymbol{V}}$, we obtain

$$\mathbb{V}[\boldsymbol{r} \mid \boldsymbol{q}] = \left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}\left(\widetilde{\boldsymbol{V}}^{-1} + \boldsymbol{X}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\boldsymbol{\Sigma}^{-1}\right)^{-1}$$

$$\mathbb{E}[\boldsymbol{r} \mid \boldsymbol{q}] = \mathbb{V}[\boldsymbol{r} \mid \boldsymbol{q}]\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\left(\widetilde{\boldsymbol{V}}^{-1} + \boldsymbol{X}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\widetilde{\boldsymbol{V}}^{-1}\widetilde{\boldsymbol{\xi}}$$

Step 6: Calculate the mean-variance optimal portfolio.

Solution:

$$w^* = \lambda^{-1}\mathbb{V}[r \mid q]^{-1}\mathbb{E}[r \mid q]$$
$$= \lambda^{-1}\Sigma^{-1}X\left(\widetilde{V}^{-1} + X^\top\Sigma^{-1}X\right)^{-1}\widetilde{V}^{-1}\widetilde{\xi}$$

8   Meucci's framework

In the realm of financial modelling and investment strategy, Meucci's framework stands out as an exceptional approach. Its superiority can be attributed to several compelling factors. Firstly, it significantly enhances our understanding of market dynamics and allocation strategies by providing profound insights into market invariants. Secondly, the framework is exhaustive, seamlessly integrating all stages of the investment process, from inception to culmination. Finally, the hallmark of Meucci's framework is its extraordinary versatility, reminiscent of a set of LEGO bricks. It offers the flexibility to effortlessly modify or substitute estimation methods or allocation techniques as required. This adaptability grants an unparalleled level of freedom in its application, further emphasizing the unique advantages of Meucci's framework.

8.1   Investor's Profile

Idea: The investor's profile consists of 3 major parts to describe the investor's intention.

1. **Target Market:** The market the investor wants to invest in. For an option trader, the market is the option market; for a retiree, the market is a set of mutual funds.

2. **Investment Horizon τ:** How long the action of investing is going to last, e.g., A day-trader aims at cashing profits within a few hours from the investment decision. A retiree has an investment horizon of the order of a few years.

3. **Investor's Objective:** The quantity $\Psi_{\boldsymbol{\alpha}}$ that an investor wants to maximize.

$$\Psi_{\boldsymbol{\alpha}} \equiv \boldsymbol{\alpha}^{\top}\mathbf{M}, \quad \mathbf{M} = \boldsymbol{a} + \mathbf{B}\mathbf{P}_{T+\tau}$$

where $\boldsymbol{a}$ is a suitable conformable vector and $\mathbf{B}$ is a suitable conformable invertible matrix.

Context: Suppose the vector of prices of n securities at a generic time $t \in \mathbb{R}$ is $\mathbf{P}_t \in \mathbb{R}^n$ (a vector of random variables), the vector of portfolio weights is $\boldsymbol{\alpha} \in \mathbb{R}^n$, the investment horizon is $\tau \in \mathbb{R}$, and the investor makes the investment decision at time $T \in \mathbb{R}$. Then, the investor's current wealth is $W_T(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^{\top}\mathbf{p}_T$ where we use lower case to emphasize the fact that at the current time T, the price for securities is known. His/Her wealth at the end of his investment horizon $T + \tau$ is $W_{T+\tau}(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^{\top}\mathbf{P}_{T+\tau}$. Then, the investor might have the following objectives $\Psi_{\boldsymbol{\alpha}}$:

- **Absolute wealth/Final wealth (For average human/retail investor):** The total amount of money you have at the end of your investment horizon.

$$\Psi_{\boldsymbol{\alpha}} \equiv W_{T+\tau}(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^{\top}\mathbf{P}_{T+\tau} = \boldsymbol{\alpha}^{\top}\mathbf{M}_{\text{absolute wealth}}$$

  where $\mathbf{M}_{\text{absolute wealth}} = \mathbf{P}_{T+\tau}, \boldsymbol{a} = \mathbf{0}, \mathbf{B} = \mathbf{I}_n$.

- **Relative wealth (For portfolio manager):** The total excess amount of money you have at the end of your investment horizon with respect to a

benchmark.

$$\Psi_{\boldsymbol{\alpha}} \equiv W_T(\boldsymbol{\alpha}) \left(\frac{W_{T+\tau}(\boldsymbol{\alpha})}{W_T(\boldsymbol{\alpha})} - \frac{W_{T+\tau}(\boldsymbol{\alpha}_{\text{benchmark}})}{W_T(\boldsymbol{\alpha}_{\text{benchmark}})} \right)$$

$$= \boldsymbol{\alpha}^\top \left(\mathbf{I}_n - \frac{\mathbf{p}_T \boldsymbol{\alpha}_{\text{benchmark}}^\top}{\boldsymbol{\alpha}_{\text{benchmark}}^\top \mathbf{p}_T} \right) \mathbf{P}_{T+\tau}$$

$$\boldsymbol{\alpha}^\top \mathbf{M}_{\text{realative wealth}}$$

where $\mathbf{M}_{\text{realative wealth}} = \left(\mathbf{I}_n - \frac{\mathbf{p}_T \boldsymbol{\alpha}_{\text{benchmark}}^\top}{\boldsymbol{\alpha}_{\text{benchmark}}^\top \mathbf{p}_T} \right) \mathbf{P}_{T+\tau}, \boldsymbol{a} = \mathbf{0}, \mathbf{B} = \left(\mathbf{I}_n - \frac{\mathbf{p}_T \boldsymbol{\alpha}_{\text{benchmark}}^\top}{\boldsymbol{\alpha}_{\text{benchmark}}^\top \mathbf{p}_T} \right).$

- **Net profits (For traders):** The profit you have at the end of your investment.

$$\Psi_{\boldsymbol{\alpha}} \equiv W_{T+\tau}(\boldsymbol{\alpha}) - w_T(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^\top \left(\mathbf{P}_{T+\tau} - \mathbf{p}_T \right) = \boldsymbol{\alpha}^\top \mathbf{M}_{\text{net profit}}$$

where $\mathbf{M}_{\text{net profit}} = \left(\mathbf{P}_{T+\tau} - \mathbf{p}_T \right), \boldsymbol{a} = -\mathbf{p}_T, \mathbf{B} = \mathbf{I}_n.$

4. **Investment constraints:** For instance, consider budget constraints: these reflect the amount of money the investor is willing to invest at the time they make their investment decisions. Normally, this also takes into account transaction costs.

Remark 8.1. We see that the objective function $\Psi_{\boldsymbol{\alpha}}$ is

- homogeneous of first degree:

$$\Psi_{\lambda\alpha} = \lambda \Psi_{\alpha}$$

- additive:

$$\Psi_{\alpha+\beta} = \Psi_{\alpha} + \Psi_{\beta}.$$

These properties allow building and **comparing** objectives that refer to complex portfolios of securities. △

8.2   Modeling the Market

Idea: The idea from Meucci's book, Risk and Asset Allocation, is that we first find an invariant $X_{t,\tau}$ for the target market, and then model this invariant using some statistical techniques on some small interval $\tilde{\tau}$ which is typically smaller than the investment horizon τ. Next, we project the estimated invariant $X_{t,\tilde{\tau}}$ on to $X_{T+\tau}$ to find the estimated invariant for the investment horizon. Then we project the estimated invariant for the investment horizon to prices of the securities $\mathbf{P}_{T+\tau}$ via some mapping. Finally, we can do whatever we want with the estimated prices.

8.2.1   Find Market's Invariants

To ascertain this, we fix an estimation interval $\tilde{\tau}$ (e.g. one week) and a starting point $\tilde{t}$ (e.g. five years ago) and we consider the set of stock prices at the equally spaced estimation times

$$P_t, \quad t \in \mathcal{D}_{\tilde{t},\tilde{\tau}}, \quad \mathcal{D}_{\tilde{t},\tilde{\tau}} = \{\tilde{t}, \tilde{t} + \tilde{\tau}, \dots, T\}$$

Each of these random variables becomes available at the respective time t.

We want the market invariant $X_{t,\tilde\tau}$ to have the following desired properties:

- $X_{t,\tilde\tau}$'s are **identically distributed.**

  We apply the **Split Test** to examine this property where you split the time series $X_{t,\tau}, t \in \mathcal{D}_{\tilde t, \tilde\tau}$ into two series:

$$X_{t,\tilde\tau}, \quad t = \tilde t, \ldots, \tilde t + \left[\frac{T - \tilde t}{2\tilde\tau}\right]\tilde\tau$$

$$X_{t,\tilde\tau}, \quad t = \left(\left[\frac{T - \tilde t}{2\tilde\tau}\right] + 1\right)\tilde\tau, \ldots, T,$$

  where $[\cdot]$ denotes the floor operator. Then we compare the respective histograms. If $X_{t,\tilde\tau}$ is an invariant, in particular all the terms in the series are identically distributed: therefore the two histograms should look very similar to each other.

- $X_{t,\tilde\tau}$'s are **independent of each other.**

  We apply the **Lag Plot Test** to examine this property where you plot x_t versus $x_{t-\tilde\tau}, \quad t = \tilde t + \tilde\tau, \ldots, T.$

  If X_t is an invariant, in particular all the terms in the series are independent of each other: therefore the scatter plot must be symmetrical with respect to the reference axes. Furthermore, since all the terms are identically distributed, the scatter plot must resemble a circular cloud.

- $X_{t,\tilde\tau}$ is **time homogeneous.** i.e. It doesn't depend on a well chosen starting time $\tilde t$.

- $X_{t,\tilde\tau}$ is **additive.** Specifically, it would be better if it is in a form of differences.

Remark 8.2. In Meucci's book, he purposed to use Elliptical distribution which is a generalization of the normal distribution, to model the distribution of the market invariant. This is too much of an overkill, we just stick to model the distribution of the market invariant by multivariate normal distribution. △

Proposition 8.3. *(Invariants for Specific Markets)*

1. *Equities, commodities, exchange rates:* $X_{t,\tilde\tau}$ *is compounded returns where the compounded return at time t for a horizon τ is defined as follows:*

$$C_{t,\tilde\tau} \equiv \ln\left(\frac{P_t}{P_{t-\tilde\tau}}\right), t \in \mathcal{D}_{\tilde t, \tilde\tau}$$

2. *Fixed-income market:* $X_{t,\tilde\tau}$ *is changes in yield to maturity*

$$X_{t,\tilde\tau}^{(v)} \equiv Y_t^{(v)} - Y_{t-\tilde\tau}^{(v)}$$

where the yield to maturity of a zero-coupon bond that expires at time $T + v$ at a generic time t is defined as follows:

$$Y_t^{(v)} \equiv -\frac{1}{v}\ln\left(Z_t^{(t+v)}\right)$$

, $Z_t^{(E)}$ *is a zero-coupon bond that a certain amount of money is turned in at the generic time* t *and a (larger) determined amount is received back at a later, specified maturity date* E. *Since the amount to be received is determined, we can normalize it as follows without loss of generality:*

$$Z_E^{(E)} \equiv 1.$$

3. **Derivatives:** $X_{t,\tilde\tau}$ *is changes in at-the-money-forward (ATMF) implied percentage volatility with generic fixed rolling expiry* v:

$$X_{t,\tilde\tau}^{(v)} \equiv \sigma_t^{(K_t,t+v)} - \sigma_{t-\tilde\tau}^{(K_{t-\tilde\tau},t-\tilde\tau+v)}, \quad t \in \mathcal{D}_{\tilde t, \tilde\tau}.$$

Each of these random variables becomes available at the respective time t.

Then, we can compress alll of the market invariants in one single vector, and denote it by $\mathbf{X}_{t,\tilde\tau} \in \mathbb{R}^d$.

8.2.2   Model the distribution of the market invariants

Context: Suppose we have our market invariant $\mathbf{X}_{t,\tilde\tau} \in \mathbb{R}^d, t \in \mathcal{D}_{\tilde t, \tilde\tau} = \{\tilde t, \tilde t + \tilde\tau, \dots, T\}$, and, by choice, they are i.i.d., time homogeneous and additive. Now, we want to estimate it. WLOG, we can reduce our notation for invariants to time series vectors

$$\mathbf{x}_t \in \mathbb{R}^d, t \in \{1, 2, 3, \dots, N\}$$

via a linear mapping on the original t where N is an integer.

8.2.3   Non-parametric estimators

These estimators are based on the law of large numbers. Therefore, they perform well when the number of empirical observations in the time series of the market invariants is large.

Proposition 8.4. *(Plug-in Estimators)*

- **Sample Mean:** $\widehat{\mathrm{E}}\,[\mathbf{x}] = \frac{1}{N}\sum_{t=1}^N \mathbf{x}_t$.

- **Sample Covariance:** $\widehat{\mathrm{Cov}}\,[\mathbf{x}] = \frac{1}{N}\sum_{t=1}^N \left(\mathbf{x}_t - \widehat{\mathrm{E}}\,[\mathbf{x}]\right)\left(\mathbf{x}_t - \widehat{\mathrm{E}}\,[\mathbf{x}]\right)^\top$

**Remark 8.5.** Not good at handling outliers.   △

Proposition 8.6. *(Kernel Density Estimators)* *The basic idea behind KDE is to smooth out the data by replacing each data point with a small kernel, or "bump," centered at that point. The kernels are typically Gaussian (bell-shaped) functions, but other shapes can be used as well. The kernels for different data points are then added together to create a continuous estimate of the probability density function* $f_{\mathbf{x}}$. *Suppose the bandwidth* $\mathbf{H} = diag(h^2, \dots, h^2) \in \mathbb{R}^{d\times d}$ *and the kernel function is*

$$K(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{1}{2}\mathbf{x}^\top \mathbf{x}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-1/2} K\left(\mathbf{H}^{-1/2}\mathbf{x}\right)$$

Then, the kernel density estimate is

$$\hat{f}_{\mathbf{x}} \mapsto \hat{f}_{\mathbf{x};\mathbf{H}} \equiv \frac{1}{N} \sum_{t=1}^{N} K_{\mathbf{H}}\left(\mathbf{x} - \mathbf{x}_t\right)$$

$$= \frac{1}{N} \sum_{t=1}^{N} \frac{1}{(2\pi)^{\frac{d}{2}} |\mathbf{H}|^{1/2}} e^{-\frac{1}{2}\left(\mathbf{H}^{-1/2}(\mathbf{x}-\mathbf{x}_t)\right)^{\top}\left(\mathbf{H}^{-1/2}(\mathbf{x}-\mathbf{x}_t)\right)}$$

$$= \frac{1}{N} \sum_{t=1}^{N} \frac{1}{(2\pi)^{\frac{d}{2}} h^d} e^{-\frac{1}{2h^2}(\mathbf{x}-\mathbf{x}_t)^{\top}(\mathbf{x}-\mathbf{x}_t)}$$

The outcome of this operation is a smoother empirical probability density function such as the one sketched in Figure 14 for $d = 2$.
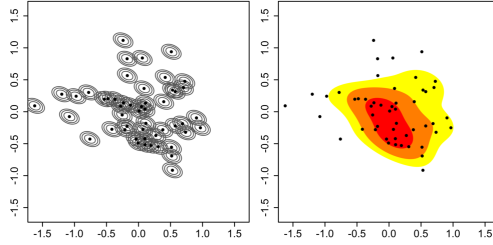
Figure 14: Construction of 2D kernel density estimate. Left. Individual kernels. Right. Kernel density estimate.

Remark 8.7.

The bandwidth of the kernel must be chosen according to the following trade-off. A narrow bandwidth overfit, so the estimator produced is not robust; a wide bandwidth underfits, so the estimator produced gives rise to loss of information. △

8.2.4   maximum likelihood estimators

MLE are built in such a way that the past observations of the market invariants, $\mathbf{x}_t \in \mathbb{R}^d, t \in \{1, 2, 3, \ldots, N\}$, become the most likely outcomes of the estimated parametric distribution.

We assume the probability density function of the market invariant $\mathbf{x}_i$ (i.i.d.) to be $f_{\boldsymbol{\theta}}$, where $\boldsymbol{\theta}$ is an S-dimensional parameter that fully determines the distribution of $\mathbf{x}_i$ and that ranges in a given set Θ.

Then, we want to maximize the likelihood function $f_{\boldsymbol{\theta}}\left(\mathbf{x}_1\right) \cdots f_{\boldsymbol{\theta}}\left(\mathbf{x}_T\right)$ of the time series, or we can maximize the log-likelihood function $\sum_{t=1}^{T} \ln f_{\boldsymbol{\theta}}\left(\mathbf{x}_t\right)$. Therefore the maximum likelihood estimator (MLE) of the parameters $\boldsymbol{\theta}$ is defined as follows:

$$\widehat{\boldsymbol{\theta}}_{MLE} \equiv \operatorname*{argmax}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \prod_{t=1}^{N} f_{\boldsymbol{\theta}}\left(\mathbf{x}_t\right) = \operatorname*{argmax}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{t=1}^{N} \ln f_{\boldsymbol{\theta}}\left(\mathbf{x}_t\right).$$

Property 8.8. *(Properties of MLE)*

(a) *invariance property the MLE of a function of the parameters is that function applied to the MLE of the parameters:*

$$\widehat{g(\boldsymbol{\theta})}_{MLE} = g(\widehat{\boldsymbol{\theta}}_{MLE}).$$

(b) *the following relation holds in approximation*

$$\widehat{\boldsymbol{\theta}}_{MLE} \sim \mathrm{N}\left(\boldsymbol{\theta}, \frac{\boldsymbol{\Gamma}}{T}\right).$$

the approximation becomes exact as T tends to infinity

In this expression $\boldsymbol{\Gamma}$ is a symmetric and positive matrix called the Fisher information matrix:

$$\boldsymbol{\Gamma} \equiv \mathrm{Cov}\left\{\frac{\partial \ln\left(f_{\boldsymbol{\theta}}(\mathbf{X})\right)}{\partial \boldsymbol{\theta}}\right\},$$

(c) *The Cramer-Rao lower bound theorem states that the inefficiency of the maximum likelihood estimator, is the smallest possible achievable with an unbiased estimator.*

Remark 8.9. When the number of observations is not very large. as it turns out, the main driver of the performance of the maximum likelihood estimators is the overall level of correlation among the market invariants, as summarized by the condition number.   △

8.2.5 Shrinkage Estimators

When the number of observations is so small, it's time to use the shrinkage estimators.

Suppose we have the sample mean and the sample covariance matrix:

$$\widehat{\mathbf{m}} \equiv \frac{1}{T}\sum_{t=1}^{T}\mathbf{c}_{t,\tilde{\tau}}, \quad \widehat{\mathbf{S}} \equiv \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{c}_{t,\tilde{\tau}} - \widehat{\mathbf{m}}\right)\left(\mathbf{c}_{t,\tilde{\tau}} - \widehat{\mathbf{m}}\right)^{\top}.$$

We then shrink the covariance matrix toward a spherical estimator:

$$\widehat{\boldsymbol{\Sigma}} \equiv (1-\epsilon)\widehat{\mathbf{S}} + \frac{\epsilon}{N}\sum_{n=1}^{N}\widehat{S}_{nn}\mathbf{I}_{N}$$

where from the shrinkage weight reads:

$$\epsilon \equiv \frac{1}{T}\frac{\frac{1}{T}\sum_{t=1}^{T}\mathrm{tr}\left\{\left(\mathbf{c}_{t,\tilde{\tau}}\mathbf{c}_{t,\tilde{\tau}}^{\top} - \widehat{\mathbf{S}}\right)^{2}\right\}}{\mathrm{tr}\left\{\left(\widehat{\mathbf{S}} - \frac{1}{N}\sum_{n=1}^{N}\widehat{S}_{nn}\mathbf{I}_{N}\right)^{2}\right\}}$$

Finally, we shrink the sample mean towards a target vector:

$$\widehat{\boldsymbol{\mu}} \equiv (1 - \gamma)\widehat{\mathbf{m}} + \gamma\mathbf{b}$$

In this expression the shrinkage target follows from:

$$\mathbf{b} \equiv \frac{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\mathbf{m}}}{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}} \mathbf{1}, \quad \gamma \equiv \frac{1}{T} \frac{\sum_{n=1}^{N} \widehat{\Sigma}_{nn} - 2\lambda_1}{(\widehat{\mathbf{m}} - \mathbf{b})^\top (\widehat{\mathbf{m}} - \mathbf{b})}$$

where λ_1 is the highest eigenvalue of $\widehat{\Sigma}$.

8.2.6   Some practical tips

There are times we need to conduct special treatment to improve the estimation of the distribution of market invariants in specific situations.

1. **Outliers:** We address outlier detection using high breakdown estimators, such as the minimum volume ellipsoid and the minimum covariance determinant.

2. **Missing Data:** We handle missing data with the help of the EM algorithm.

3. **Adjusting for economical understandings:** We use weighted estimation techniques like exponential smoothing, which takes into account the greater reliability of more recent data compared to older data.

8.3   Project the estimated invariants to the investment horizon

Idea: Basically, we want to find the distribution of $\boldsymbol{X}_{T+\tau,t}$ which is determined by $\phi_{\boldsymbol{X}_{T+\tau,t}}$ using the distribution of $\boldsymbol{X}_{T+\tilde{\tau},\tilde{t}}$ which is determined by $\phi_{\boldsymbol{X}_{T+\tilde{\tau},\tilde{t}}}$.

$$\boldsymbol{X}_{T+\tilde{\tau},\tilde{t}} \mapsto \boldsymbol{X}_{T+\tau,t}$$

With our fine choice of invariants, this is actually very easy to accomplish.

Since all of our invariants are specifically chosen to be in the form of differences, they are additive, i.e. they satisfy the following relation:

$$\mathbf{X}_{T+\tau,\tau} = \mathbf{X}_{T+\tau,\tilde{\tau}} + \mathbf{X}_{T+\tau-\tilde{\tau},\tilde{\tau}} + \cdots + \mathbf{X}_{T+\tilde{\tau},\tilde{\tau}}.$$

Also, since the terms in the right-hand side of the equation above are invariants relative to non-overlapping time intervals, they are , by choice, independent and identically distributed random variables. Therefore, we must have

$$\phi_{\mathbf{X}_{T+\tau,\tau}} = \left(\phi_{\mathbf{X}_{t,\tilde{\tau}}}\right)^{\frac{\tau}{\tilde{\tau}}} \iff \mathbf{X}_{T+\tau,\tau} \sim \mathcal{N}\left(\frac{\tau}{\tilde{\tau}}\widehat{\boldsymbol{\mu}}, \frac{\tau}{\tilde{\tau}}\widehat{\boldsymbol{\Sigma}}\right)$$

Remark 8.10. We conclude by pointing out that the simplicity of the projection formula hides the dangers of **estimation risk.**   △

8.4   Map invariants to Prices

Since we are dealing with raw securities, the pricing function has the following simple form:

$$P_{T+\tau} = P_T e^X = e^{\mathbf{Y}}$$

where the ancillary variable $\mathbf{Y}$ is an affine transformation of the market invariants:

$$\mathbf{Y} \equiv \gamma + \mathrm{diag}(\varepsilon)\mathbf{X}$$

where $\gamma_n \equiv \ln(P_T)$ and $\varepsilon_n \equiv 1$.

8.5   Evaluating Allocation

Suppose we have an efficient frontier computed by the MVO framework. The question then becomes, how should we select the optimal allocation from all the options available on the efficient frontier? To aid decision-making, Meucci proposed a metric known as the "index of satisfaction." Generally, we use **Certainty-Equivalent** as satisfaction index:

Certainty-Equivalent: Here, we select a utility function and its associated risk-aversion coefficient. Assume we have an objective Ψ_α. Then, we have

$$\boldsymbol{\alpha}^* = \max_{\boldsymbol{\alpha}} \mathrm{CE}(\boldsymbol{\alpha}) \equiv \max_{\boldsymbol{\alpha}} u^{-1}\left(\mathrm{E}\left\{u\left(\boldsymbol{\alpha}'\mathbf{M}\right)\right\}\right)$$

8.6   Mean-Variance Optimization

Motivation: The mean-variance framework can be viewed as an approximation of the more general framework, which involves the general index of satisfaction. This broader approach imposes significant computational challenges, making the mean-variance framework a more practical alternative.

The investor's index of satisfaction is a functional of the distribution of the investor's objective. Since the distribution of the investor's objective is in general uniquely determined by its moments, the index of satisfaction can be re-written as a function defined on the infinite-dimensional space of the moments of the distribution of the objective:

$$\mathcal{S}(\boldsymbol{\alpha}) \equiv \mathcal{H}\left(\mathrm{E}\left\{\Psi_\alpha\right\}, \mathrm{CM}_2\left\{\Psi_\alpha\right\}, \mathrm{CM}_3\left\{\Psi_\alpha\right\}, \ldots\right) \approx \tilde{\mathcal{H}}\left(\mathrm{E}\left\{\Psi_\alpha\right\}, \mathrm{Var}\left\{\Psi_\alpha\right\}\right).$$

for a suitable bivariate function $\tilde{\mathcal{H}}$, and CM_k denotes as in the central moment of order k of a univariate distribution, $\mathrm{CM}_k\{\Psi\} \equiv \mathrm{E}\left\{(\Psi - \mathrm{E}\{\Psi\})^k\right\}$.

Then, we first solve

$$\boldsymbol{\alpha}(v) \equiv \underset{\substack{\alpha \in \mathcal{C} \\ \mathrm{Var}\{\Psi_\alpha\}=v}}{\mathrm{argmax}} \ \mathrm{E}\left\{\Psi_\alpha\right\}$$

where $v \geq 0$ to get the efficient frontier. Then find the optimal allocation following the one-dimensional search:

$$\boldsymbol{\alpha}^* \equiv \boldsymbol{\alpha}\left(v^*\right) \equiv \underset{v \geq 0}{\mathrm{argmax}}\, \mathcal{S}(\boldsymbol{\alpha}(v)).$$

Remark 8.11. I find Meucci's framework to be superior for several compelling reasons.

1. The insights it provides regarding market invariants enhance our comprehension of the market dynamics and allocation strategies.

2. The framework is comprehensive, encompassing every step of the investment process from inception to culmination.

3. Meucci's framework is exceptionally versatile. This flexibility is akin to a set of LEGO bricks, allowing for easy alterations and substitutions in estimation methods or allocation techniques as desired. This adaptability provides unparalleled freedom in its application.

$\triangle$

9   Empirical Example

Now, let's delve into an empirical example. In this case, we use Meucci's framework as the foundation, and compare the performance of the MVO framework in various scenarios, as well as the effectiveness of the Black-Litterman APT model. Our findings suggest that to achieve a better PnL, improving the estimation of critical variables is paramount. By "better", I don't just mean statistically superior, but also more economically logical. Additionally, we demonstrated that the Black-Litterman-APT model is not the holy grain. It is only as efficient as its inputs.

9.1   Investment Profile

We denote as $\mathbf{P}_t$ the prices at the generic time t of one share of the assets. Now, we initialize the necessary ingradients for our investment.

- We assume investor's initial wealth is a given amount of cash $W = 1000\$$.

- We assume that the investor chooses a set of $N \equiv 7$ stocks and that the investor plans to re-invest any dividends.

- We assume that the investment horizon is $\tau = 3$ weeks.

- We assume that the investor's objective is final wealth:

$$\Psi_\alpha \equiv \boldsymbol{\alpha}^\top \mathbf{P}_{T+\tau}$$

  where T denotes the time investor makes his/her investment decision.

- We assume the investor cares about the certainty-equivalent of his expected utility where his utility function is of the power type.

$$u(x) = x^{\frac{\gamma-1}{\gamma}}$$

  Therefore the investor's satisfaction reads:

$$\mathcal{S}(\boldsymbol{\alpha}) = CE \equiv \left(\mathbb{E}\left\{ \Psi_\alpha^{\frac{\gamma-1}{\gamma}} \right\} \right)^{\frac{\gamma}{\gamma-1}}$$

  We pick $\gamma = -9$ for the risk aversion parameter of the power utility functions.

Remark 9.1.

1. Notice that the power utility function is defined only for positive values of the investor's objective. This is consistent with the fact that prices are positive and that the investor can only hold long positions in the stocks.

$\triangle$

9.2 Modeling the Market

- **Market Data:**

  - The time-series of the asset's weekly prices for tickers ['AAPL', 'AMZN', 'JPM', 'JNJ', 'XOM', 'GE', 'HD'] from $2012 - 01 - 09$ to $2022 - 12 - 26$.

  - Market capitalization data for tickers ['AAPL', 'AMZN', 'JPM', 'JNJ', 'XOM', 'GE', 'HD'] from $2012 - 01 - 09$ to $2022 - 12 - 26$.

  - The time-series of the asset prices for the S&P 500 index from $2012 - 01 - 09$ to $2022 - 12 - 26$.

  All of the data can be obtained from the WRDS database and Yahoo Finance API.

- **Market Invariants:** Non-overlapping compounded returns can be modeled as independent and identically distributed across time:

$$\mathbf{X}_{t,\widetilde{\tau}}^{(n)} \equiv \ln\left(\frac{P_t^{(n)}}{P_{t-\widetilde{\tau}}^{(n)}}\right)$$

- **Estimation horizon :** $\widetilde{\tau} = 1$ week.

- Assume:

$$\mathbf{X}_{t,\tilde{\tau}} \sim \mathrm{N}\left(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}\right)$$

- We account for transaction costs in the market. In this case we assume that the transaction costs grow quadratically with the number of shares to be invested:

$$\mathcal{T}(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^{\top}\mathbf{D}\boldsymbol{\alpha}$$

  where $\mathbf{D} =$ is a diagonal matrix of positive entries. Specifically, we restrict all of its entries to be 0.001.

  Remark 9.2. The non-linear growth of the transaction costs accounts for the market impact of large stock transactions. $\triangle$

9.3 Estimators

1. **Method 1: Sample Estimator** Sample mean and the sample covariance matrix:

$$\widehat{\mathbf{m}} \equiv \frac{1}{T}\sum_{t=1}^{T}\mathbf{c}_{t,\tilde{\tau}}, \quad \widehat{\mathbf{S}} \equiv \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{c}_{t,\tilde{\tau}} - \widehat{\mathbf{m}}\right)\left(\mathbf{c}_{t,\tilde{\tau}} - \widehat{\mathbf{m}}\right)^{\top}.$$

2. **Method 2: Shrinkage Estimator** We shrink the covariance matrix toward a spherical estimator:

$$\widehat{\boldsymbol{\Sigma}} \equiv (1 - \epsilon)\widehat{\mathbf{S}} + \frac{\epsilon}{N}\sum_{n=1}^{N}\widehat{S}_{nn}\mathbf{I}_N$$

where from the shrinkage weight reads:

$$\epsilon \equiv \frac{1}{T} \frac{\frac{1}{T}\sum_{t=1}^{T} \text{tr}\left\{\left(\mathbf{c}_{t,\tilde{\tau}}\mathbf{c}_{t,\tilde{\tau}}^{\top} - \widehat{\mathbf{S}}\right)^2\right\}}{\text{tr}\left\{\left(\widehat{\mathbf{S}} - \frac{1}{N}\sum_{n=1}^{N} \widehat{S}_{nn}\mathbf{I}_N\right)^2\right\}}$$

Finally, we shrink the sample mean towards a target vector:

$$\widehat{\boldsymbol{\mu}} \equiv (1-\gamma)\widehat{\mathbf{m}} + \gamma\mathbf{b}$$

In this expression the shrinkage target follows from:

$$\mathbf{b} \equiv \frac{\mathbf{1}^{\top}\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{m}}}{\mathbf{1}^{\top}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}\mathbf{1}, \quad \gamma \equiv \frac{1}{T}\frac{\sum_{n=1}^{N}\widehat{\Sigma}_{nn} - 2\lambda_1}{(\widehat{\mathbf{m}} - \mathbf{b})^{\top}(\widehat{\mathbf{m}} - \mathbf{b})}$$

where λ_1 is the highest eigenvalue of $\widehat{\Sigma}$.

3. **Method 3: Black-Litterman-APT**

   As we stated in Section 6.2, we choose market betas, size, volatility, and momentum, a total of four common risk premia as the elements in the $\boldsymbol{f}$ process in the APT model.

   $$\boldsymbol{r}_{t+\tilde{\tau}} = \mathbf{X}_t\boldsymbol{f}_t + \boldsymbol{\epsilon}_t, \quad \mathbb{E}[\boldsymbol{\epsilon}] = 0, \quad \mathbb{V}[\boldsymbol{\epsilon}] = \mathbf{D}$$

   We designate the first four columns of $\mathbf{X}$ to represent exposures to the four risk premia. For each estimation period, denoted as $[t, t + \tilde{\tau}]$, we can calculate the exposure as follows:

   (a) **Market beta:** Each asset's daily excess return time series is regressed against the S&P 500 excess return time series, over the full scale estimating dataset or a trailing 30 days window.

   (b) **Size:** Market capitalization.

   (c) **Volatility:** Similar to the market beta calculation, the daily excess return time series for each asset is regressed against the S&P 500 excess return time series. This is done either over the full-scale estimation dataset or a trailing 30-day window. The mean-square error (MSE) resulting from this regression is considered the raw exposure.

   (d) **Momentum:** Compound returns over the full scale estimating dataset or a trailing 30 days window.

   Remark 9.3. Since the number of assets for our example is pretty small, we don't have to winsorize and gaussianize them.   △

   Then, firstly, we calculate the OLS factor returns as $\hat{\boldsymbol{f}}_t = \left(\boldsymbol{X}_t'\boldsymbol{X}_t\right)^{-1}\boldsymbol{X}_t'\boldsymbol{r}_{t+\tilde{\tau}}$.

   We repeat this process over the dataset for intervals $[t, t + \tilde{\tau}], [t, t + 2\tilde{\tau}], \ldots, [t, T]$, resulting in estimations for the $\boldsymbol{f}$ process and its corresponding residual $\boldsymbol{\epsilon}$.

   We then use the mean and covariance of our estimations for $\hat{\boldsymbol{f}}_t$ to represent the mean, as denoted by $\boldsymbol{\xi}$, and variance $\mathbf{V}$ in our data-driven prior, with $\pi(\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\xi}, \mathbf{V})$, and we consider the covariance of the corresponding residuals $\boldsymbol{\epsilon}$ as $\mathbf{D}$.

Next, as outlined in **Step 4:** Specify the view equation., we assume the investor's views are focused solely on risk factors, and he thinks the risk factors will be as good as the risk factors at the time of making investment decisions T. Hence, we define $q = \hat{f}_T$ and set the confidence matrix Ω of our views to $\mathbf{I}q \cdot 0.0009$, where $\mathbf{I}q$ is an identity matrix.

Lastly, we proceed to find $\mathbb{V}[r \mid q]$ and $\mathbb{E}[r \mid q]$ as stated in **Step 5**, treating them as the variance and mean for the invariants $\mathbf{X}$. This allows us to proceed within Meucci's framework.

9.4 Project the distribution of the invariants to the investment horizon

We project the distribution of the invariants to the investment horizon and obtain that compounded returns from the investment date to the investment horizon are normally distributed:

$$\mathbf{X}_{T+\tau,\tau} \sim \mathcal{N}\left(\frac{\tau}{\tilde{\tau}}\widehat{\boldsymbol{\mu}}, \frac{\tau}{\tilde{\tau}}\widehat{\boldsymbol{\Sigma}}\right) \equiv \mathcal{N}\left(\widehat{\boldsymbol{\mu}}^{\mathbf{X}}, \widehat{\boldsymbol{\Sigma}}^{\mathbf{X}}\right)$$

Since in the context of equity market, $\mathbf{P}_{T+\tau,\tau} \sim logN\left(\frac{\tau}{\tilde{\tau}}\widehat{\boldsymbol{\mu}}, \frac{\tau}{\tilde{\tau}}\widehat{\boldsymbol{\Sigma}}\right)$, the mean and variance of $\mathbf{P}_{T+\tau,\tau}$ are given by

$$\mathbb{E}\left\{P_{T+\tau}^{(n)}\right\} = P_T^{(n)} e^{\widehat{\boldsymbol{\mu}}_n^{\mathbf{X}} + \frac{\widehat{\boldsymbol{\Sigma}}_{nn}^{\mathbf{X}}}{2}}$$

and

$$\mathrm{Cov}\left\{P_{T+\tau}^{(m)}, P_{T+\tau}^{(n)}\right\} = P_T^{(m)} P_T^{(n)} e^{\widehat{\boldsymbol{\mu}}_m^{\mathbf{X}} + \widehat{\boldsymbol{\mu}}_n^{\mathbf{X}} + \frac{1}{2}\left(\widehat{\boldsymbol{\Sigma}}_{mm}^{\mathbf{X}} + \widehat{\boldsymbol{\Sigma}}_{nn}^{\mathbf{X}}\right)}\left(e^{\widehat{\boldsymbol{\Sigma}}_{mn}^{\mathbf{X}}} - 1\right)$$

9.5 Computing the optimal allocation

Step 1: Specify constraints. We formulate the investor's constraints as

- **budget constraint:**
$$\mathcal{C}_1 : \boldsymbol{\alpha}^\top \mathbf{p}_T \leq W_T - \boldsymbol{\alpha}^\top \mathbf{D}\boldsymbol{\alpha}$$

where W_T is his capital at the time where the investor makes his investment decisions.

- **Long-only Constraint:**
$$\mathcal{C}_2 : \boldsymbol{\alpha} \geq \mathbf{0}$$

Step 2: Construct the MVO framework. Fitting the constraints into the mean-variance framework, we have

$$\boldsymbol{\alpha}^{(i)} \equiv \arg\max_{\boldsymbol{\alpha}} \boldsymbol{\alpha}^\top \mathbb{E}\left\{\mathbf{P}_{T+\tau}\right\}$$

$$\text{subject to} \begin{cases} \boldsymbol{\alpha}^\top \mathrm{Cov}\left\{\mathbf{P}_{T+\tau}\right\}\boldsymbol{\alpha} \leq v^{(i)} \\ \boldsymbol{\alpha}^\top \mathbf{p}_T \leq w - \boldsymbol{\alpha}^\top \mathbf{D}\boldsymbol{\alpha} \\ \boldsymbol{\alpha} \geq \mathbf{0}. \end{cases}$$

Step 3: Compute the efficient frontier. We choose a grid of $I \equiv 100$ target variances $\{v^{(1)}, \ldots, v^{(I)}\}$ and solve numerically each time the above optimization.

Each optimization is a quadratically constrained linear programming problem which is a subproblem of the Second-Order Cone Programming (SOCP) problem, so it can be easily solved using the "CVXOPT" package.

Remark 9.4.

1. In fitting the framework of SOCP solver, we need to decompose the covariance $\text{Cov}\{\mathbf{P}_{T+\tau}\} = \mathbf{L}\mathbf{L}^T$ using Cholesky decomposition technique and transform the constraint to
$$\|\mathbf{L}^\top \boldsymbol{\alpha}\| \leq v^{(i)}$$

   Same applies to the budget constraint.

2. Do not set $v^{(1)}$, the smallest variance, to zero as that is unrealistic.

$\triangle$

Step 4: Determine the optimal allocation.

Since the investor uses the index of satisfaction as a metric for evaluating allocations, we can determine the optimal allocation accordingly as the optimal allocation is the portfolio that gives rise to the higher level of satisfaction.

To determine this portfolio we use Monte Carlo simulations. We simulate a large number J of Monte Carlo market scenarios as follows:
$$_j\mathbf{P}_{T+\tau} \equiv \text{diag}(\mathbf{p}_T) \, e^j \mathbf{C}$$

where the exponential acts component-wise and where each vector $_j\mathbf{C}$ is an independent drawing from the multivariate normal distribution for all $j = 1, \ldots, J$. In our example we perform $J = 100 + 2i$ where i is the number of decisions the investor has made just for convenience for all of the simulations except for the one model with sample estimator we use a fixed $J = 100$.

We compute the following approximation for all the mean-variance efficient portfolios in the grid:
$$\widetilde{\mathcal{S}}\left(\boldsymbol{\alpha}^{(i)}\right) \equiv \left(\frac{1}{J}\sum_{j=1}^{J}\left(_j\mathbf{P}'_{T+\tau}\boldsymbol{\alpha}^{(i)}\right)^{\frac{\gamma-1}{\gamma}}\right)^{\frac{\gamma}{\gamma-1}}$$

We rank the levels of satisfaction provided by the mean-variance efficient portfolios:
$$i^* \equiv \underset{i}{\text{argmax}}\left\{\widetilde{\mathcal{S}}\left(\boldsymbol{\alpha}^{(i)}\right)\right\}$$

Finally, we determine the optimal allocation:
$$\boldsymbol{\alpha}^* \equiv \boldsymbol{\alpha}^{(i^*)}$$

9.6   Models Summary

Here, we briefly summarize the five models under analysis. It's important to note that all five models use compound returns as invariants, absolute wealth as objectives, a one-week estimation horizon, a three-week investment horizon, and a certainty-equivalent of expected utility as the index of satisfaction, with the utility function being of the power type. The models only differ in the methods used to estimate the mean and variance of the invariants, the datasets used for estimation, and the number of iterations in Monte Carlo we use to find the optimal allocation on the efficient frontier.

1. **Model 1:** In this model, we use a sample estimator over the full scale estimating dataset with a fixed number of iterations, $J = 100$, to find the optimal allocation.

2. **Model 2:** In this model, we use a sample estimator over the full scale estimating dataset with an increasing number of iterations, $J = 100 + 2i$, to find the optimal allocation, where i is the number of investment decisions made.

3. **Model 3:** In this model, we use a shrinkage estimator over the full scale estimating dataset with an increasing number of iterations, $J = 100 + 2i$, to find the optimal allocation, where i is the number of investment decisions made.

4. **Model 4:** In this model, we use a Black-Litterman-APT estimator over the full scale estimating dataset with an increasing number of iterations, $J = 100 + 2i$, to find the optimal allocation, where i is the number of investment decisions made.

5. **Model 5:** In this model, we use a Black-Litterman-APT estimator over a trailing 30 weeks window with an increasing number of iterations, $J = 100 + 2i$, to find the optimal allocation, where i is the number of investment decisions made.

Remark 9.5. We've chosen i to represent the number of investment decisions made. This choice doesn't carry particular significance but serves as a convenient way to increase the number of iterations in line with capital increases.   △

9.7   Plots and Analysis

Analysis 1: Distortion introduced by additional constraints.

We first examine the MVO framework in a single period for **Model 1**. We get to the conclusion that the efficient frontier is distorted by the various constraints we impose on the optimization process, and it's easy for the mean-variance framework to generate corner solutions.

We have the following results:

In the plot on the left, referenced as Figure 15, we illustrate the allocation weights in terms of asset units. On the other hand, in the plot on the right, referenced as Figure 16, we present the allocation weights as relative portfolio weights.

It's clearly evident that as the allowed variance (risk) increases, the allocation on the efficient frontier tends to converge towards a corner solution. This outcome should be
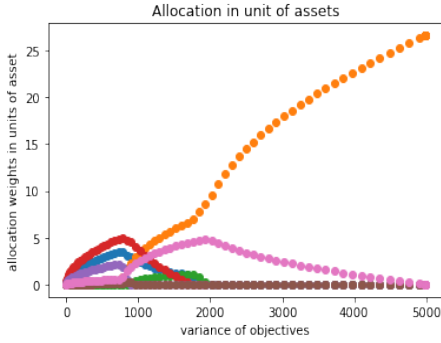
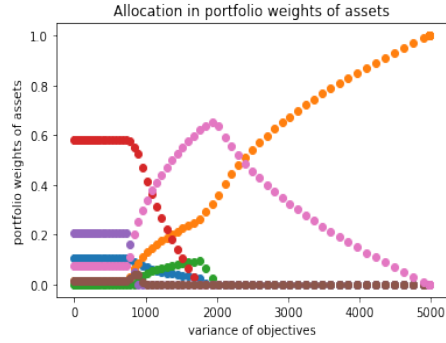**Figure 15:** Allocation weights in units of asset(Single Period)

Figure 16: Allocation weights (Single Period)

expected since we imposed additional constraints on the classical MVO framework (e.g. long-only). Such optimization problems often have a high likelihood of finding their maximizers/minimizers at the boundary.
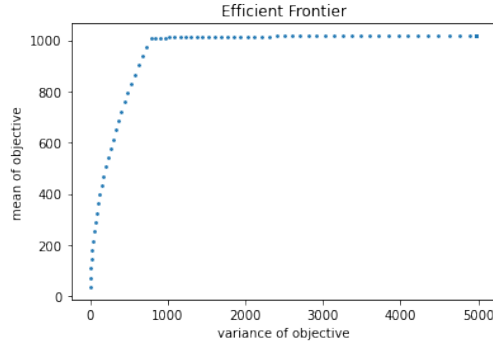

Figure 17: Single Period Allocation Efficient Frontier

In Figure 17, we display the risk/reward profile of the mean-variance efficient allocations in terms of expected value and variance of final wealth:

$$\mathrm{E}\left\{\Psi_{\alpha}\right\} = \boldsymbol{\alpha}^{\top}\mathrm{E}\left\{\mathbf{P}_{T+\tau}\right\}, \quad \mathrm{Var}\left\{\Psi_{\alpha}\right\} = \boldsymbol{\alpha}^{\top}\mathrm{Cov}\left\{\mathbf{P}_{T+\tau}\right\}\boldsymbol{\alpha}.$$

It's easy to see that this efficient frontier looks different from the one we generated using fictional data in the section about classical MVO. This is because the various constraints we impose on the optimization process distort the behavior of the optimal allocation, resulting in a steep increase at first which then becomes flat afterwards.

Analysis 2: APT Model

Here, we analyze the four risk factors as stated in Item 3 estimated over the full scale dataset from $2012 - 01 - 09$ to $2017 - 01 - 02$.

Since the APT model is essentially a linear model, whether the risk factors serve as positive or negative drivers depends solely on their associated coefficients in $\boldsymbol{f}$. The cumulative factor return is simply the cumulative sum of all $\hat{\boldsymbol{f}}_t$'s over a period of time.
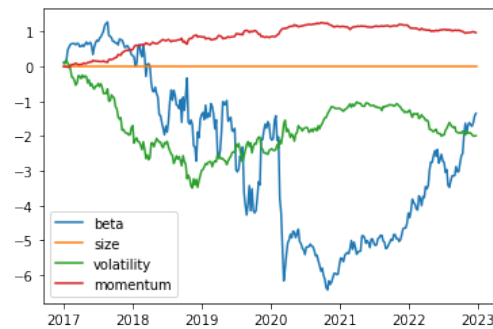
Figure 18: Cumulative factor returns to risk premia $\hat{f}_t$

For each entry in the cumulative factor returns, if the plot shows a downward trend over time, it suggests that the most recent entries are turning negative. This indicates that there is a minus sign before the corresponding risk premia during that period, which implies that the risk premia serve as a negative driver for the stock price, and vice versa.

Figure 18 is consistent with our intuition. The cumulative factor returns to momentum risk premia (i.e. coefficients before the momentum risk premia) is consistently positive, meaning the sign before the momentum risk premia is mostly positive. This suggests that, in most cases, a stock with higher momentum tends to outperform one with lower momentum, after controlling for other systematic sources of risk. The cumulative factor returns to volatility risk premium is negative, meaning the sign before the volatility risk premia is mostly negative. This suggests that, in most cases, stocks with higher volatility tend to perform poorly after controlling for other systematic sources of risk. The size risk premium is flat, likely because all chosen stocks have very high market capitalization, rendering the influence of size negligible when comparing the performance of these selected stocks.

The weird shape of the market beta factor return actually makes sense. The downward trend of the market beta entry in the cumulative factor return during 2017-2019 suggests that the coefficients for market beta were negative most of the time during this period, implying that stocks correlating well with the market were punished. Since all our chosen stocks are from large companies, which usually correlate well with the market, most of them under-performed during this period. The drastic drop in the spring of 2020 due to COVID-19 could provide further evidence for this, as it indicates a big big negative sign before the market beta risk premium at the time. Even stocks with only slight correlation with the market were significantly, negatively impacted. However, at the end of 2020, as Biden started his rescue plan, the markets began to recover and the market beta entry in the cumulative factor return f started to increase. This suggests that the most recent coefficients (the market beta entries of the factor return f) for market betas were turning positive, implying that stocks correlating well with the market were beginning to perform well. We can see from Figure 19 and Figure 20 below that it was indeed the case.
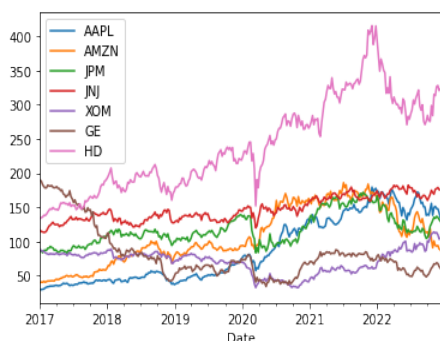
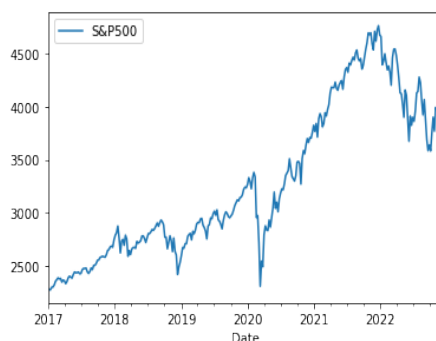**Figure 19:** Chosen stocks performance

Figure 20: Market performance

Analysis 3: Models Comparison

In this analysis, we compare the PnL curve of the five models that we purposed as shown in Figure 21.

1. **Finding 1: Shrinkage estimator perform better than vanilla sample estimator.**

   When comparing the performance of the model that uses the sample estimator with the one that uses the shrinkage estimator, it's easy to see that the shrinkage estimator outperforms the sample estimator consistently. This makes sense as the shrinkage estimator trades a bit more bias for less variance, resulting in a more robust and stable estimate. However, it's important to note that the impact of the estimators is relatively small. The most significant differences between the models still lie in the way they are constructed.
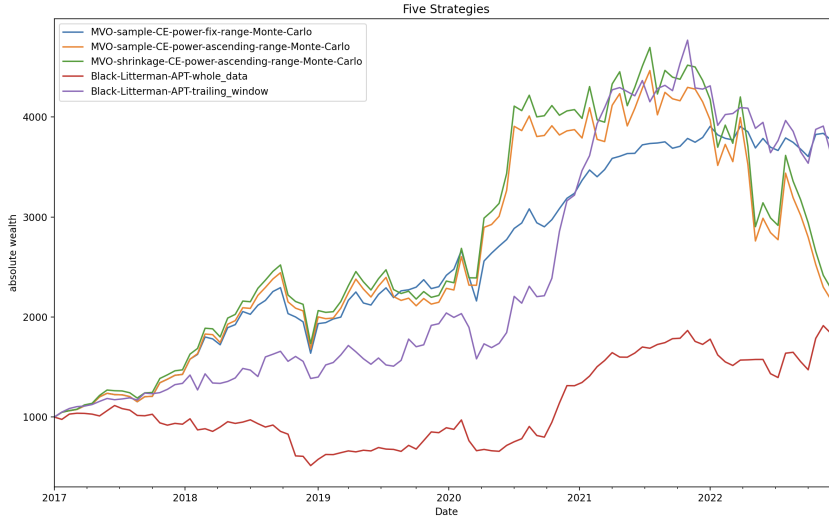
2. **Finding 2: Maximizing certainty-equivalent does NOT mean maximizing absolute wealth.**

   When comparing the performance of the MVO model with an increasing upper bound for the concerning variance to that of the MVO model with a fixed upper bound for the concerning variance, we observe that while the former generally outperforms the latter, it fails to maintain its lead towards the end. This suggests that optimal decisions guided by the certainty-equivalent index of satisfaction do not always guarantee superior performance. Essentially, this information implied by the index of satisfaction instructs you on how to make optimal investment decisions according to your risk preference.

3. **Finding 3: Adjusting for economic understanding is essential.**

   When comparing the performance of the Black-Litterman-APT model with estimations conducted on a full-scale dataset — where the mean of the f -process is significantly influenced by data far removed from the time when the investor makes their decision — and that of the Black-Litterman-APT model with estimations based on a trailing 30-weeks window, it's apparent that the model that adjusts for economic understanding, or what one might call data significance,

Figure 21: Performance of the five models proposed

performs far better. This has led to the development of several practical techniques used by practitioners for data transformation, aiming to account for the greater reliability of more recent data compared to older data, such as exponential decay/smoothing.

4. **Finding 4: Black-Litterman-APT model is not the holy grain.**

Unlike the classical Black-Litterman model, which anchors its return to the market portfolio, the Black-Litterman-APT model does not anchor to anything. Alternatively, one could say its return is pegged on the inputs fed into the model. As in **Finding 3**, the performance of the Black-Litterman-APT model with estimations conducted on a full-scale dataset — where the mean of the f-process is significantly influenced by data far removed from the time when the investor makes their decision — is way worse than that of the Black-Litterman-APT model with estimations based on a trailing 30-weeks window. This suggests that the performance of the Black-Litterman-APT model is dictated by how good your view q is and how good your estimations are for the f-process. However, the performance of the Black-Litterman-APT model is indeed much more stable.

10   Future Work: MVO with transaction cost

The traditional Mean-Variance Optimization (MVO) approach often results in portfolio allocations that incur substantial trading costs, significantly impacting realized risk-adjusted returns. As demonstrated in the empirical example, we incorporated transaction costs into the classical MVO framework; however, this model is simplified and does not fully capture the complexities of reality.

In the real world, transaction costs include direct costs such as commissions and taxes, the bid-ask spread, and indirect costs like slippage and market impact. Slippage is defined as the price difference between the time a trade is anticipated and the volume-weighted average price at which it executes. Market impact refers to the price changes caused by the trade itself, which can be substantial when the size-to-average volume ratio of trades is high.

Fortunately, several models have been proposed in the literature to address these issues, such as those by Hasbrouck (1991), Lillo, Farmer, and Mantegna (2003), and Almgren, Thum, Hauptmann, and Li (2005). Consequently, in future research, we could consider incorporating their ideas into our framework.

11 Appendix

11.1 Von Neumann-Morgenstern Utility Theory

In decision theory, the von Neumann – Morgenstern (VNM) utility theorem shows that any individual whose preferences satisfied four VNM axioms has a utility function (VNM Theorem); such an individual's preferences can be represented by $U(\mathbf{p})$, the von Neumann – Morgenstern utility function. And the individual will always prefer actions that maximize expected utility (Expected Utility Hypothesis). The theorem is the basis for expected utility theory.

11.1.1 VNM – Formulate the Question

VNM start with Knightian Risk (finite probability space) and assumed:

(1) **Outcomes:** mutually exclusive outcomes $(s_1, s_2, \ldots, s_n)$.

    E.g. Toss of a coin – two possible outcomes; heads (s_1) or tails (s_2).

(2) **Prizes:** The monetary amounts that each outcome maps to.

(3) **Lottery:** A scenario where a set of probabilities (Probability Vector) $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ is assigned to the outcomes $(s_1, s_2, \ldots, s_n)$.

(4) **Lottery Set:** Define $\Delta(s)$ as the set of all lotteries on s; that is, the set of all probability vectors $\mathbf{p} \in \mathbb{R}^n$.

    Clearly $\Delta(s)$ is convex; if $p_1, p_2 \in \Delta(s)$, then for any $0 \leq \alpha \leq 1$, $\alpha p_1 + (1 - \alpha)p_2 \in \Delta(s)$ must hold.

(5) **Preference Function:** A ranking between lotteries in $\Delta(s)$ is a binary relation $\succeq$.

    (a) For two lotteries $\mathbf{p}, \mathbf{q} \in \Delta(s)$ we write $\mathbf{p} \succeq \mathbf{q}$ if $\mathbf{p}$ is preferred to or equivalent to $\mathbf{q}$.

    (b) Equivalence is denoted by $\equiv$ and means both $\mathbf{p} \succeq \mathbf{q}$ and $\mathbf{q} \succeq \mathbf{p}$.

    (c) $\mathbf{p} \succ \mathbf{q}$ means strictly preferred.

(6) Economic agent ranks lotteries (probability vectors).

Ø **Example:** 3 outcomes $\longrightarrow 3$ prizes:

$$s_1 = \text{you receive } \$0$$
$$s_2 = \text{you receive } \$500,000,000$$
$$s_3 = \text{you receive } \$1,000,000,000$$

People decide between lotteries $\mathbf{p}_1 = (0, 1, 0)$ and $\mathbf{p}_2 = (.5, 0, .5)$.

11.1.2   VNM Axioms

(1) **Transitivity:** For $\mathbf{p}, \mathbf{q}, \mathbf{r} \in \Delta(s)$, $\mathbf{p} \succeq \mathbf{q}$ and $\mathbf{q} \succeq \mathbf{r}$ means $\mathbf{p} \succeq \mathbf{r}$.

$\implies$ No circular preferences: if I like chocolate better than vanilla and vanilla better than strawberry, then I like chocolate better than strawberry.

(2) **Completeness:** For $\mathbf{p}, \mathbf{q} \in \Delta(s)$, either $\mathbf{p} \succeq \mathbf{q}$ or $\mathbf{q} \succeq \mathbf{p}$ or both. If both are true then we say $\mathbf{p} \equiv \mathbf{q}$.

$\implies$ Only one view.

$\implies$ Can be lotteries that are not the same but are equivalent.

$\implies$ **Reflexivity:** ($\mathbf{p} \succeq \mathbf{p}$ so $\mathbf{p} \equiv \mathbf{p}$) follows from Completeness.

(3) **Independence:** If $\mathbf{p}, \mathbf{q}, \mathbf{r} \in \Delta(s)$ with $\mathbf{p} \succeq \mathbf{q}$ and $\alpha \in [0, 1]$, then $\alpha\mathbf{p} + (1 - \alpha)\mathbf{r} \succeq \alpha\mathbf{q} + (1 - \alpha)\mathbf{r}$.

$\implies$ If I like chocolate better than strawberry, then I like (a mixture of chocolate and vanilla) better than (the same mixture of strawberry and vanilla).

$\implies$ **Strong/Debatable Axiom**: there could be interactions between items that would alter the results.

(4) **Continuity:** If $\mathbf{p}, \mathbf{q}, \mathbf{r} \in \Delta(s)$ with $\mathbf{p} \succeq \mathbf{q} \succeq \mathbf{r}$, then there is a scalar $\alpha \in [0, 1]$ such that $\alpha\mathbf{p} + (1 - \alpha)\mathbf{r} = \mathbf{q}$.

$\implies$ If I prefer chocolate to vanilla, and I prefer vanilla to strawberry, then there is some mixture of chocolate and strawberry that I like the same as vanilla.

$\implies$ **Strong/Debatable Axiom**: I might hate the taste of mixtures and dislike even the smallest adulteration of my beloved chocolate, so there is no combination that gets me to the same satisfaction level as vanilla.

Remark 11.1. Eventually these axioms go bad - people don't actually behave like this if we push it far enough. $\triangle$

11.1.3   VNM Theorem

VNM Theorem quantified the more intuitive preference functions by showing that

$$\text{utility functions} \quad = \quad \text{preference functions.}$$

Theorem 11.2. *(VNM Theorem) let $\Delta(s)$ be a convex subset of $\mathbb{R}^n$. Let $\succeq$ be a binary relation on $\Delta(s)$. Then $\succeq$ satisfies the four VNM axioms if and only if there is a function $U : \Delta(s) \to \mathbb{R}$ such that:*

(a) *U represents $\succeq$. (i.e. $\forall \mathbf{p}, \mathbf{q} \in \Delta(s), \mathbf{p} \succeq \mathbf{q} \iff U(\mathbf{p}) \geq U(\mathbf{q})$)*

(b) *U is affine. (i.e. $\forall \mathbf{p}, \mathbf{q} \in \Delta(s), \alpha \in [0, 1], U(\alpha\mathbf{p} + (1 - \alpha)\mathbf{q}) = \alpha U(\mathbf{p}) + (1 - \alpha)U(\mathbf{q})$)*

Solution: Set elementary lotteries $\mathbf{x}_i = \mathbf{e}_i$ where the ith outcome s_i has 100% probability of happening. Suppose $\mathbf{x}_n \succeq \mathbf{x}_{n-1} \succeq \cdots \succeq \mathbf{x}_1$. Start construction of the utility function U by defining $U(\mathbf{x}_1) = u(s_1) = 0$ and $U(\mathbf{x}_n) = u(s_n) = 1$, unless $x_n \equiv x_1$ in which case the trivial utility function $U(\mathbf{x}) = 0$ for all lotteries $\mathbf{x}$.

By the continuity axiom, there is an α_i so that $\alpha_i \mathbf{x}_n + (1 - \alpha_i)\mathbf{x}_1 \equiv x_i$ for any $1 \le i \le n$. Then set $U(\mathbf{x}_i) = u(s_i) = \alpha_i$. Then, for any lottery $\mathbf{p} = (p_1, \ldots, p_n) = \sum_{i=1}^n \mathbf{x}_i p_i$, we have

$$U(\mathbf{p}) = p_1 U(\mathbf{x}_1) + p_2 U(\mathbf{x}_2) + \cdots + p_n U(\mathbf{x}_n)$$
$$= p_1 u(s_1) + p_2 u(s_2) + \cdots + p_n u(s_n)$$

where u assigns to each outcome s_i a real number $u(s_i)$.

This construction makes sense because for any lottery $\mathbf{p}$

$$\mathbf{p} = (p_1, \ldots, p_n) = \sum_{i=1}^n \mathbf{x}_i p_i$$
$$\equiv \sum_{i=1}^n (\alpha_i \mathbf{x}_n + (1 - \alpha_i)\mathbf{x}_1)p_i$$
$$= \left(\sum_{i=1}^n \alpha_i p_i \right) \mathbf{x}_n + \left(1 - \sum_{i=1}^n \alpha_i p_i \right) \mathbf{x}_1$$
$$= U(\mathbf{p})\mathbf{x}_n + (1 - U(\mathbf{p}))\mathbf{x}_1$$

That is, lottery $\mathbf{p} = (p_1, \ldots, p_n) \equiv (1 - U(\mathbf{p}), 0, \ldots, 0, U(\mathbf{p}))$.

It follows that

$$\mathbf{p} \succeq \mathbf{q}$$
$$\iff (1 - U(\mathbf{p}), 0, \ldots, 0, U(\mathbf{p})) \succeq (1 - U(\mathbf{q}), 0, \ldots, 0, U(\mathbf{q}))$$
$$\text{By Lemma} \iff U(\mathbf{p}) \ge U(\mathbf{q})$$

which is consistent with our aim of construction. Also, $U(\mathbf{p})$ is apparently affine.

Remark 11.3.

1. Utility Functions are here to help us rank lotteries, so there is no true meaning of them.

2. U is unique up to a positive linear transformation: If $V : \Delta(s) \to \mathbb{R}$ also satisfies (a) and (b), then there are $b, c \in \mathbb{R}$ (where $b > 0$) such that $V = bU + c$.

3. Normally, we will use this theorem in the reverse order by directly saying a person's utility function is xxx so that we have a preference relation.

4. $U(\mathbf{p})$ is called the VNM utility functions, and $u(s_i)$ is called the Bernoulli utility functions.

5. **Non-Satiation:** All else being constant, individuals always prefer more of positive goods rather than negative goods, vice versa. Therefore, we have $u^\top(x) \geq 0$.

$\triangle$

Theorem 11.4. *(Expected Utility Hypothesis) An individual chooses not the highest expected value, but rather the highest expected utility; that is one would want to maximize*

$$U(\mathbf{p}) = p_1 U(\mathbf{x}_1) + p_2 U(\mathbf{x}_2) + \cdots + p_n U(\mathbf{x}_n)$$
$$= p_1 u(s_1) + p_2 u(s_2) + \cdots + p_n u(s_n)$$

Definition 11.5. *(Certainty Equivalent) Given a utility function on prizes and a lottery $\mathbf{p}$, the certainty equivalent of $\mathbf{p}$ is the number m such that*

$$u(m) = \mathbb{E}_{\mathbf{p}}[u(X)]$$

where m is sure prize that has the same utility as risky outcomes encoded in X with probabilities in $\mathbf{p}$.

11.1.4   Problems with utility theory

Why? 1.the overstrong set of assumptions about what constitutes rational decision making, 2. humans are irrational.

Problem 11.6. (The Allais Paradox)

(1) Decision 1: Choose between two lotteries. $\mathbf{p}_1$ gives a 63% chance of $\$500,000 \cdot \mathbf{p}_2$ gives a 61% chance of $\$520,000$.

(2) Decision 2: Choose between two lotteries. $\mathbf{q}_1$ gives a 100% chance of $\$500,000.\mathbf{q}_2$ gives a 98% chance of $\$520,000$.

Should be clear that $\mathbf{p}_2, \mathbf{q}_1$ (which is fairly common) violates utility theory.

Solution: Most people choose 1a and 2b. But if 1a is preferred to 1b, we must have

$$.61u(520) + .39u(0) > .63u(500) + .37u(0)$$

where u(x) means the utility of gaining x thousand dollars. Thus

$$.61(u(520) - u(500)) > .02(u(500) - u(0))$$

Assuming only that the utility function u is sensible (more is preferred to less, $u^\top > 0$), both sides of this last inequality are positive.

Preferring the certain 2b to the highly probable 2a means

$$u(500) > .98u(520) + .02u(0) > 0$$

which gives

$$.02(u(500) - u(0)) > .98(u(520) - u(500))$$

Thus for any sensible utility function, $1a \succ 1b$ and $2b \succ 2a$ are incompatible.

Conclusion: We will see various laws indicating that relying on any mathematical finance or economics "law" too heavily will break. The point of financial and economic modeling is to inform human intuition, not (yet) to replace it. Utility remains a powerful concept and a good first-order guide to economic decision-making.

References

[1] Meucci, A. *Risk and asset allocation*. Springer Science & Business Media, 2009.

[2] Black, F., & Litterman, R. B. *The Journal of Fixed Income*, 1(2), 7-18. https://-doi.org/10.3905/jfi.1991.408013, 1991.

[3] Luenberger, D. G. *Investment Science*. New York: Oxford University Press, 1998.

[4] Henderson, H. V., & Searle, S. R. *On Deriving the Inverse of a Sum of Matrices*. SIAM Review, 23(1), 53 − 60. https://doi.org/10.1137/1023004, 1981.

[5] He, G., & Litterman, R. *The Intuition Behind Black-Litterman Model Portfolios*. Available at SSRN: https://ssrn.com/abstract=334304 or http://dx.doi.org/10.2139/ssrn.334304.

[6] Kolm, P. N., Focardi, S. M., & Fabozzi, F. J. *Incorporating Trading Strategies in the Black-Litterman Framework*. In F. J. Fabozzi (Ed.), Handbook of Finance. https://-doi.org/10.1002/9780470404324.hof002036, 2008.

[7] Ritter, G. *Stable Linear-Time Optimization in Arbitrage Pricing Theory Models*. Risk Magazine. Available at SSRN: https://ssrn.com/abstract=2821360, 2016.

[8] Kolm, P., & Ritter, G. *On the Bayesian Interpretation of Black − Litterman*. European Journal of Operational Research, 258(2), 564 − 72. https://-doi.org/10.1016/j.ejor.2016.10.027, 2017.

[9] Theil, H. and Goldberger, A. S. *On pure and mixed statistical estimation in econometrics*, International Economic Review, vol. 2, no. 2, pp. 65–78, 1961.