Fundamentals of Optimization and Application to Portfolio Selection

CQF

In this lecture...

I. Introduction to optimization:

- how to formulate an optimization problem;
- elementary rules and tips.

II. Unconstrained optimization problems

- how to use calculus to solve unconstrained optimization problems;
- application to linear regression;
- application to mean-variance optimization.

III. Optimization problems with equality constraints

- the method of Lagrange;
- application to portfolio selection;
- the minimum variance portfolio;
- the tangency portfolio.

V. A short note on optimization problems with inequality constraints: the Kuhn-Tucker conditions.
VI. Extending Mean-Variance Portfolio Optimziation to benchmarks and active portfolio management.

Part I: Introduction to Optimization

An optimization problem is one in which you are trying to find the "best possible value" that a function, say f, can take subject to a number of constraints. This generally involves finding the minimum or maximum of f or of a function built around f.

Optimization techniques are very often used in finance to solve a wide array of problems ranging from calculating the value of a bond yield, to solving a portfolio selection problem in discrete and continuous time and to valuing derivatives such as American and passport options.

As optimization problems come in all shape and forms, you should expect some to be incredibly easy to solve and some other to be incredibly hard. In any case, formulating the problem well will spare you many headaches.

Formulating an optimization problem

Optimization problems are most often defined as

$$\min_{x_1,\ldots,x_n} f(x_1,\ldots,x_n) \tag{1}$$

subject to:

$$\left.\begin{array}{c} g_1(x_1,\ldots,x_n) \\ \vdots \\ g_m(x_1,\ldots,x_n) \end{array}\right\} \begin{array}{c} \leq \\ = \\ \geq \end{array} \left\{\begin{array}{c} b_1 \\ \vdots \\ b_m \end{array}\right.$$

The function f is called the **objective function**. It is the function we want to optimize (here minimize).

The variables $x_1, \dots, x_n$ are the **decision variables** with respect to which we want to optimize the function.

The functions $g_1, \dots, g_m$ are the m **constraints** faced in our optimization. These constraints can be equalities or inequalities.

Elementary rules & tips

The following set of rules will help you manipulate your optimization problem to put it in a more convenient formulation.

- some problems are more easily dealt with as a minimization, some others as a maximization. To change between the two, remember that

$$\max f(x) = -\min(-f(x))$$

$$\min f(x) = -\max(-f(x))$$

- **affine transformation** multiplying your objective function by a positive constant b and adding a constant a leaves the optimal value of the decision variable x unchanged, but remember to reverse the transformation at the end of the problem to get the correct value for the objective function

$$\max_x(a + bf(x)) = a + b\max_x(f(x)) \qquad b > 0$$

  The same is true with the min function.
- if instead of minimizing or maximizing a function, you want to set it to a value c, then you need to minimize the distance (or norm) of $f - c$. This is, however, beyond the scope of this presentation.

Part II: Unconstrained optimization

In the absence of any constraint, our system (1) reduces to:

$$\min_{x_1,\ldots,x_n} f(x_1,\ldots,x_n)$$

Now, this is just standard (albeit multivariate) calculus!

So we know that the minimal value of f is an extremum of the function.

How can we pick a vector $x^* = (x_1^*,\ldots,x_n^*)$ so that f reaches a global minimum?

From standard calculus, we know that:

- the gradient (vector of derivatives) of f at x^*, denoted by $\nabla f(x^*)$, must be zero. This is a **necessary condition**, but it is not sufficient as minima, maxima and inflection points all have a derivative reaching 0!
- the Hessian (i.e. matrix of second derivatives) of f at x^*, denoted by $Hf(x^*)$, must be positive definite (negative definite for a maximization). This is a **sufficient condition**.

These two conditions are fundamental in optimization. The first condition is used to find a set of potential solutions and the second is used to check which of these answers satisfy(ies) the problem.

They are referred to as first order (necessary) condition and second order (sufficient) condition.

Application 1 - The mean-variance optimisation criterion

We start by setting the stage.

We are in an economy with n different assets. Each asset i is entirely characterized by its expected return μ_i and expected standard deviation σ_i. In addition, assets i and j are correlated with correlation ρ_{ij}. The proportion of the portfolio invested in asset i is w_i.

The vector of asset expected returns μ is defined as:

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_i \\ \vdots \\ \mu_n \end{pmatrix}$$

The covariance matrix Σ is given by:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & & & \\ \rho_{n1}\sigma_n\sigma_1 & \cdots & \cdots & \sigma_n^2 \end{pmatrix}$$

The vector of weights is:

$$w = \begin{pmatrix} w_1 \\ \vdots \\ w_i \\ \vdots \\ w_n \end{pmatrix}$$

Therefore, the portfolio return μ_π can be written as

$$\mu_\pi = \mu' w$$

and the portfolio variance, σ_π^2, is given by

$$\sigma_\pi^2 = w' \Sigma w$$

where the superscript $'$ denotes the transpose of the vector.

Aside: A Useful Covariance Matrix Decomposition

If you have the correlation matrix and the standard deviation vector, you can calculate the covariance matrix in just two steps.

Step 1: Form the diagonal standard deviation matrix S

Consider the standard deviation vector σ:

$$\sigma = \begin{pmatrix} \sigma_1 \\ \vdots \\ \sigma_i \\ \vdots \\ \sigma_n \end{pmatrix}$$

As its name indicates, the diagonal standard deviation matrix is a matrix with standard deviations on its diagonal and 0 everywhere else:

$$S = D(\sigma) = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & \cdots & \sigma_n \end{pmatrix}$$

Step 2: compute the covariance matrix Σ

Consider the Correlation matrix R defined as:

$$R = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & & & \\ \rho_{n1} & \cdots & \cdots & 1 \end{pmatrix}$$

The covariance matrix Σ can be obtained by pre and post multiplying the correlation matrix by the diagonal standard deviation matrix:

$$\Sigma = SRS = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & & & \\ \rho_{n1}\sigma_n\sigma_1 & \cdots & \cdots & \sigma_n^2 \end{pmatrix}$$

Note that we should normally have $S'RS$, but S is a diagonal matrix and therefore $S' = S$.

This decomposition of the covariance matrix is not only useful for portfolio selection for risk measurement. It is actually used by RiskMetrics in the calculation of parametric Value at Risk.

(End of Aside).

Now, we introduce a risk-free asset, such as a bank account, with a certain return r, variance equal to 0 and correlation with the other (risky) assets equal to 0.

The weight w_0 of the risk-free asset is defined as the weight of the portfolio that remains once the allocation to risky assets is complete.

Mathematically,

$$w_0 = 1 - w'\mathbf{1}$$

The return of the portfolio is given by:

$$\mu_\pi = w'\mu + r(1 - w'\mathbf{1}) = r + w'(\mu - r\mathbf{1}) \tag{2}$$

So the introduction of a risk-free asset split returns between the risk-free rate r and the vector of risk premia $(\mu - r\mathbf{1})$.

One of the easiest portfolio optimisation problems also turns out to be the best model in theory: the mean-variance optimisation criterion.

The mean-variance optimisation criterion, which was first proposed by Markowitz, corresponds to finding the asset allocation w that maximizes the expected return of the portfolio subject to a penalty related to the variance of portfolio returns:

$$\max_{w} \mu_P - \frac{\lambda}{2}\sigma_P^2$$

Here, $\lambda > 0$ represents the degree of risk aversion.

As a result, we neither need a risk constraint nor a return contraint. Also, the budget constraint has vanished because the residual of the wealth not invested in risky assets is now in the risk-free asset.

To solve the optimisation problem, we need to express the previous equation as an explicit function of w. In a market with N risky assets and a risk-free asset, we get:

$$\max_{w} V(w) = \left[r + w'(\mu - r\mathbf{1})\right] - \frac{\lambda}{2} w'\Sigma w$$

This optimisation problem is particularly easy: the objective function $w'(\mu - r\mathbf{1}) - \frac{\lambda}{2} w'\Sigma w$ is quadratic in the decision variable w and there is no constraint.

The first order condition is

$$\nabla V(w^*) = \frac{\partial V}{\partial w}(w^*) = (\mu - r\mathbf{1}) - \lambda\Sigma w^* = 0 \tag{3}$$

which gives us the candidate solution

$$w^* = \frac{1}{\lambda}\Sigma^{-1}(\mu - r\mathbf{1}) \tag{4}$$

Before concluding, we check the second order to ensure that we found the unique maximiser:

$$HV(w^*) = -\lambda\Sigma < 0 \tag{5}$$

which confirms that we have found a maximum.

Application 2 - Ordinary Least Squares Linear Regression

The Capital Asset Pricing Model (CAPM) is a fundamental result in finance theory. The basic idea of the CAPM is that the expected risk premium of any asset should be proportional to the expected risk premium of the underlying market:

$$\mathbb{E}\left[R^A - r\right] = \beta \mathbb{E}\left[R^M - r\right] \tag{6}$$

where

- R^A is the return of asset A;
- r is the risk-free rate;
- β is the degree of exposure to systematic risk;
- R^M is the return of the overall financial market.

We can rewrite the CAPM in a more familiar form as

$$\mathbb{E}\left[R_A\right] = r + \beta \mathbb{E}\left[R_M - r\right] \tag{7}$$

The CAPM is expressed in terms of expectation...

- ... which is great for forward looking estimates;
- ... but no so much for the kind of historical analysis required to estimate parameters.

In fact we do not know the slope coefficient β, so we need to estimate it based on historical data. Typically, we will estimate the β of a stock using:

- 5 years of monthly data (60 observations) for an established company;
- 2 years of weekly data (104 observations) for a fast-changing company.

In finance textbooks, the slope coefficient β of a stock is generally estimated through an **ordinary least square (OLS) linear regression**.

Now, the first step is to convert the CAPM into a model that we can apply to *historical data*. In essence, this brings us back to Sharpe's linear factor model!

Assuming we know the value for β, that our data are measured perfectly and are stationary, the 'true' model for datapoint $i = 1, \ldots, n$ will be:

$$R_i^A - r = \beta \left(R_i^M - r \right) \tag{8}$$

Note that

- in theory r is a constant so it should be the same for all the i's. However, formulating our estimation in terms of risk premia leaves the door open to using an actual TBill or bond rate which we know is not constant;
- β is the same for all the i's.

However,

- we do not know the true value of β and we will need to rely on some estimate $\hat{\beta}$;
- the data may not be entirely reliable (think of stalled quotes for small stocks).

To make the model more realistic and account for measurement errors (such as stalled quotes), we introduce a disturbance term or **error term** ϵ

$$R_i^A - r = \beta \left(R_i^M - r \right) + \epsilon_i \tag{9}$$

For mathematical reasons, the error terms ϵ_i should be IID and satisfy the following properties:

- The mean of the error term is 0: $\mathbb{E}\left[\epsilon_i\right] = 0$ for all i's;
- The variance of the error term is finite: $\mathbb{E}\left[\epsilon_i^2\right] = s^2$ for all i's;
- The covariance of the error term is 0: $\mathbb{E}\left[\epsilon_i \epsilon_j\right] = 0$ for all i and j;

Note that we do not require the error terms to be normally distributed, but is often convenient to do so.

Let's say we want to estimate the β based on n observation.

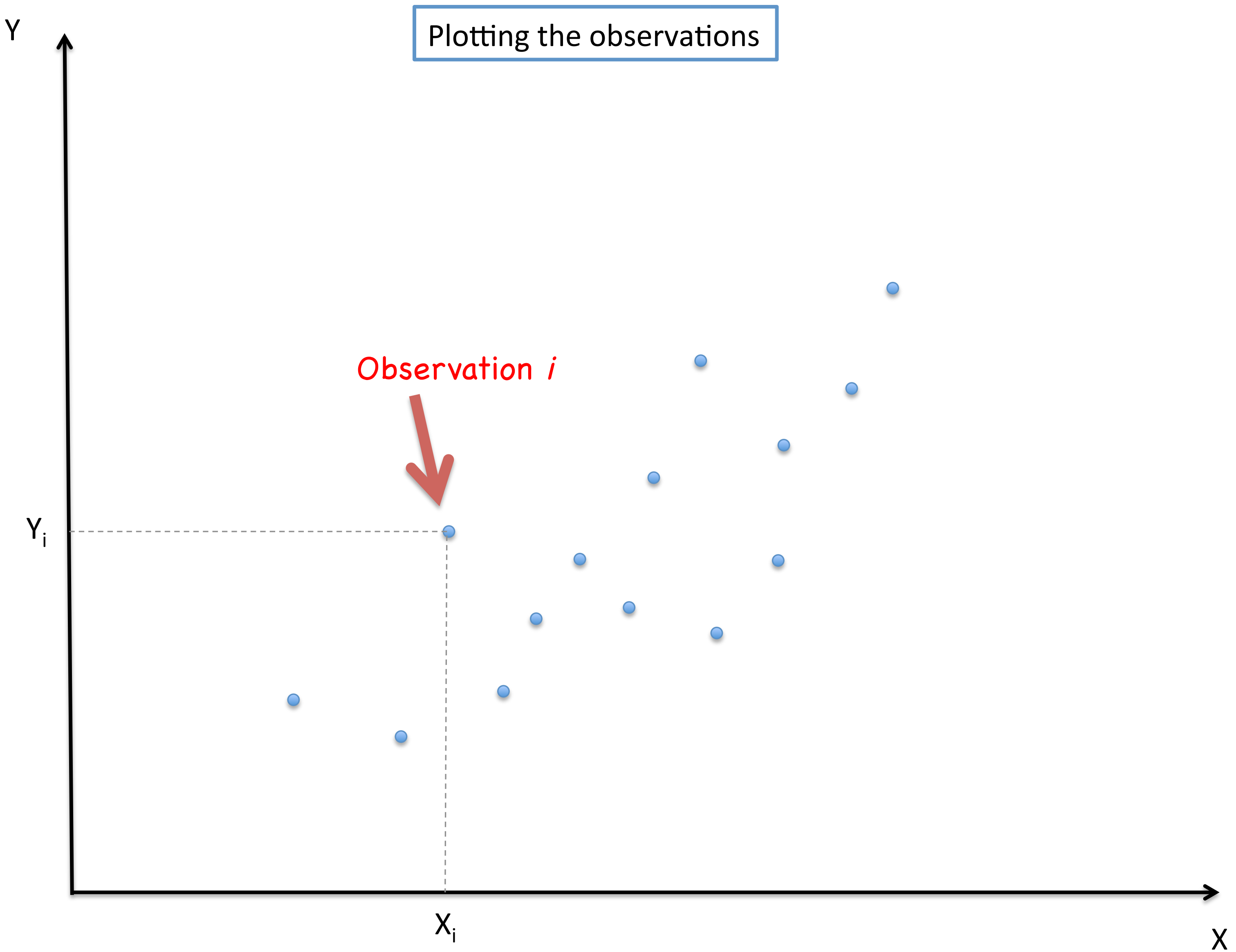
For more generality, we will solve this problem using matrix notation.

- The **dependent variable** Y is the n-element column vector of observed asset risk premia $(R^A - r)$;

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_i \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} R_1^A - r \\ \vdots \\ R_i^A - r \\ \vdots \\ R_n^A - r \end{pmatrix}$$

- The **independent variable** X is the n-element column vector of observed market risk premia $(R^M - r)$;

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} R_1^M - r \\ \vdots \\ R_i^M - r \\ \vdots \\ R_n^M - r \end{pmatrix}$$

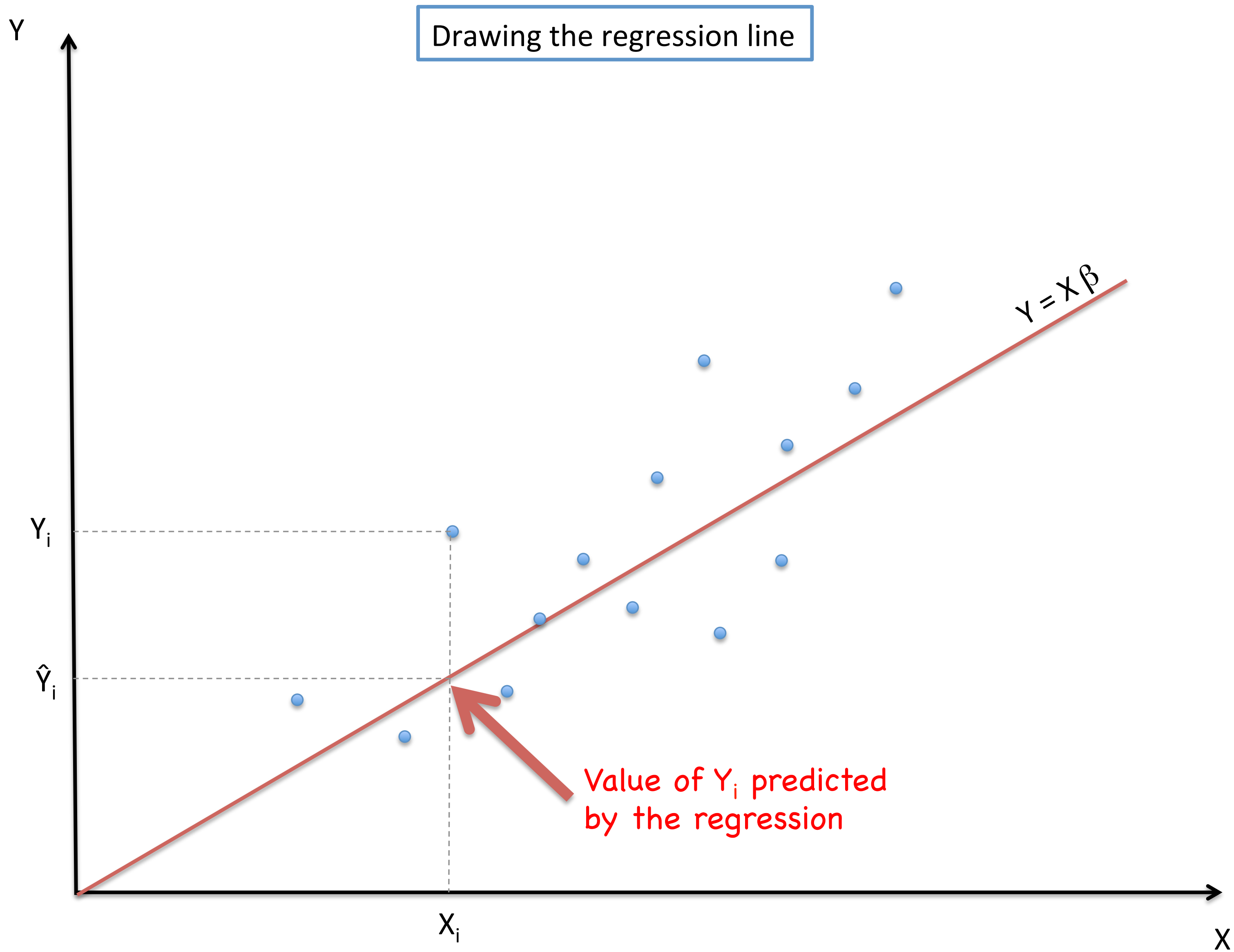

Plotting the observations
Y
Observation i
Yi
Xi
X

Thus, the model we are trying to estimate is

$$Y = X\beta + \epsilon \tag{10}$$

where ϵ is a n-element column vector of error terms:

$$\epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_n \end{pmatrix}$$



Drawing the regression line
Y
Y = X β
Y_i
$\hat{Y}_i$
Value of Y_i predicted
by the regression
X_i
X

The idea behind OLS regression is that our best estimate $\hat{\beta}$ of β must *minimize* the squared error (also called L^2 loss function in statistical estimation theory and in supervised machine learning) between the values Y we actually observe and the prediction of the linear model, that is $X\beta$:

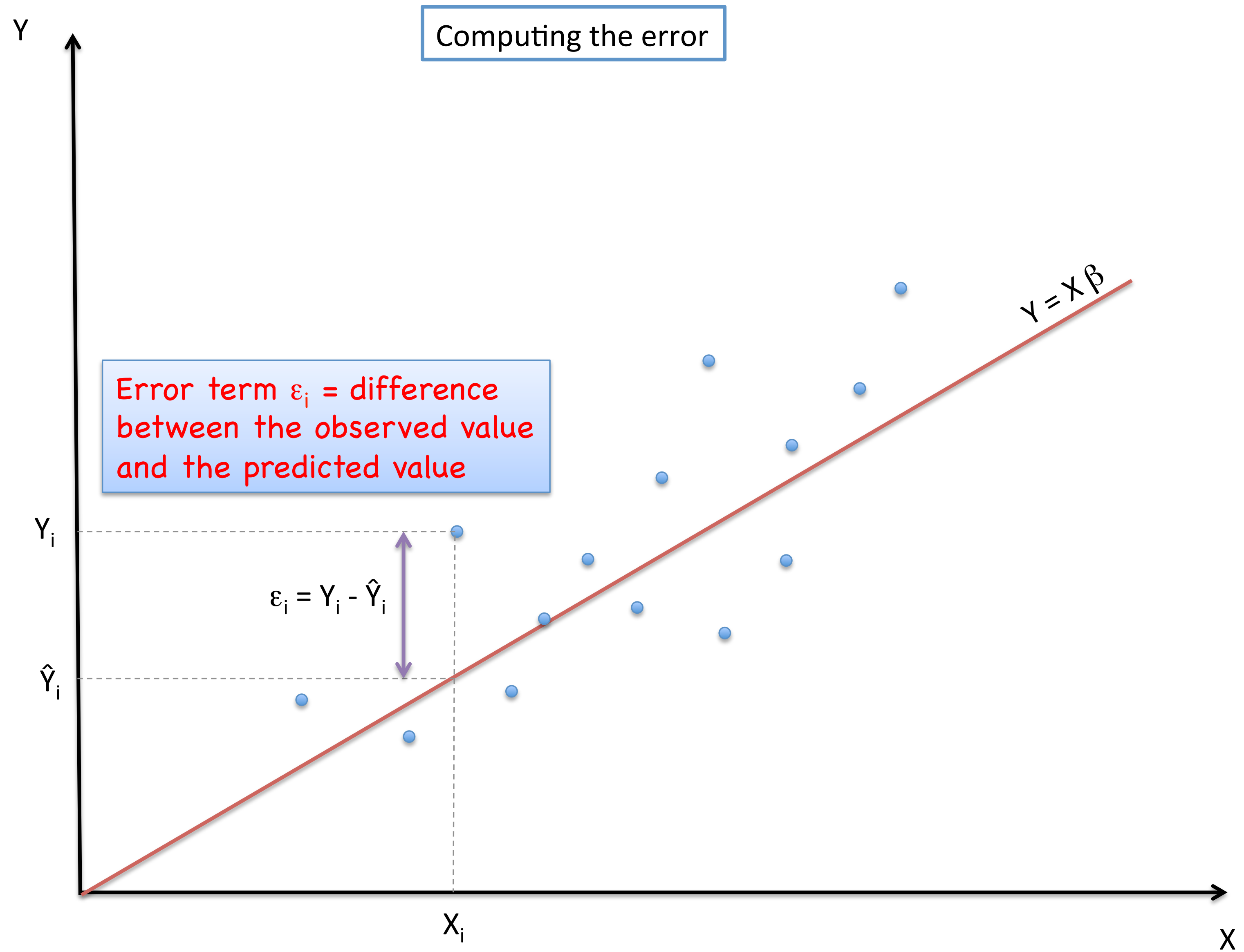
$$= \sum_{i=1}^{n} (Y_i - X_i\beta)^2 = \sum_{i=1}^{n} \epsilon_i^2 \tag{11}$$

So, this statistical estimation problem is effectively an **unconstrained optimization problem**.

By the way, why do we minimize the sum of squared error?

- the error is the signed distance between the value predicted by the model and the actual observation;
- the sum measures the total error accumulated through the n observations;
- the square prevents negative errors from offsetting positive errors while giving more weight to large errors and less to small errors.

As you will see in supervised machine learning, other loss functions exist, leading to other solutions for the statistical estimation problem.


Computing the error
Y
X
Y = X β
Error term ε_i = difference between the observed value and the predicted value
Y_i
$\hat{Y}_i$
$\varepsilon_i = Y_i - \hat{Y}_i$
X_i

In matrix notation, the problem (11) corresponds to

$$\min_{\beta} F(\beta) = (Y - X\beta)'(Y - X\beta) = \epsilon'\epsilon \tag{12}$$

where the superscript $'$ denotes the transpose of the vector.

Developing the product,

$$F(\beta) = Y'Y - 2\beta'X'Y + \beta'X'X\beta$$

The first order condition implies that

$$\frac{\partial F}{\partial \beta} = 0$$

that is

$$-2X'Y + 2X'X\beta = 0$$

and we identify the candidate solution $\hat{\beta}$

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Checking the second order condition, we deduce that $\hat{\beta}$ is the unique minimizer as long as $X'X > 0$

How good is the estimate?

The estimate $\hat{\beta}$ is actually a random variable! Except that the more data we have, the closer it gets to the true value β which is a constant.

The covariance of $\hat{\beta}$ is

$$\mathbb{E}\left[\left(\beta - \hat{\beta}\right)\left(\beta - \hat{\beta}\right)'\right] \tag{13}$$

Developping,

$$
\begin{aligned}
& \mathbb{E}\left[\left(\beta-\hat{\beta}\right)\left(\beta-\hat{\beta}\right)'\right] \\
= & \mathbb{E}\left[\left(\beta-(X'X)^{-1}X'Y\right)\left(\beta-(X'X)^{-1}X'Y\right)'\right] \\
= & \mathbb{E}\left[\left(\beta-(X'X)^{-1}X'\left(X\beta+\epsilon\right)\right)\left(\beta-(X'X)^{-1}X'\left(X\beta+\epsilon\right)\right)'\right] \\
= & \mathbb{E}\left[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}\right]
\end{aligned}
$$

The $X's$ are known, so we do not need to worry about them.

$$\begin{aligned} & \mathbb{E}\left[\left(\beta-\hat{\beta}\right)\left(\beta-\hat{\beta}\right)'\right] \\ = \quad & (X'X)^{-1}X'\mathbb{E}\left[\epsilon\epsilon'\right]X(X'X)^{-1} \end{aligned}$$

However the error vector ϵ is random. Since all the elements of this vector have same variance and are uncorrelated, the covariance matrix of ϵ is simply

$$\mathbb{E}\left[\epsilon\epsilon'\right] = s^2 I$$

where I is the identity matrix.

To conclude, the variance of the estimate $\hat{\beta}$ is

$$\mathbb{E}\left[\left(\beta-\hat{\beta}\right)\left(\beta-\hat{\beta}\right)'\right] = s^2(X'X)^{-1} \tag{14}$$

Using our linear regression function to predict Y

The value of Y predicted by the model (10), denoted by $\hat{Y}$, is

$$\hat{Y} = \mathbb{E}\left[X\hat{\beta} + \epsilon\right] = X\hat{\beta}. \tag{15}$$

Generalizing...

The matrix notation has two advantages:

- we can solve the problem very quickly;
- we have an immediate generalization to multiple linear regression!

Consider the general multifacor model

$$\tilde{R}^A = \alpha + \sum_{j=1}^{m} \beta_j \times F^j + \epsilon \tag{16}$$

where

- α is the interecept term;
- we have m factors $F^1, \dots, F^m$ and therefore m slope coefficients $\beta^1, \dots, \beta^m$

For this type of models, we have $m + 1$ parameters to estimate:

- one intercept term α ;
- m slope coefficients $\beta^1, \dots, \beta^m$

How is what we have done before going to help???

Let's say we have n historical observations for the asset returns and each factors.

We can still rewrite (16) as

$$Y = X\beta + \epsilon \tag{17}$$

BUT

- Y is the n-element column vector of observed asset return R_A

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_i \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} R_1^A \\ \vdots \\ R_i^A \\ \vdots \\ R_n^A \end{pmatrix}$$

- X is a $n \times (m+1)$ *matrix* where columns 1 to m contain the observed factor levels $F^1, \dots, F^m$ and column $(m+1)$ is the n-element unit vector (a vector of 'ones');

$$X = \begin{pmatrix} F_1^1 & \dots & F_1^j & \dots & F_1^m & 1 \\ \vdots & & & & & \\ F_i^1 & \dots & F_i^j & \dots & F_i^m & 1 \\ \vdots & & & & & \\ F_n^1 & \dots & F_n^j & \dots & F_n^m & 1 \end{pmatrix}$$

- β is a $(m+1)$-element *column* vector. The first m entries correspond respectively to $\beta^1, \ldots, \beta^m$ and the last entry is the intercept term α :

$$\beta = \begin{pmatrix} \beta^1 \\ \vdots \\ \beta^i \\ \vdots \\ \beta^m \\ \alpha \end{pmatrix} \tag{18}$$

The estimated vector β is still equal to

$$\hat{\beta} = (X'X)^{-1}X'Y$$

with covariance matrix

$$\mathbb{E}\left[\left(\beta - \hat{\beta}\right)\left(\beta - \hat{\beta}\right)'\right] = s^2(X'X)^{-1}$$

Application 3 - Generalised Least Squares Linear Regression

Generalised Least Squares (GLS) is a "simple" generalisation of ordinary least squares that is often helpful in addressing the problems of heteroskedasticity and autocorrelation of residuals.

- It was proposed in 1934 by Alexander Aitken.

Unlike OLS, GLS does not assume that

- The residuals are uncorrelated with each other, $\mathbb{E}[\epsilon_i \epsilon_j] = 0, \ i \neq j$
- The independent variable is uncorrelated with the residuals $\mathbb{E}[X_{ij}\epsilon_i] = 0, \ j = 1, \ldots, p, i = 1, \ldots, n$

Instead, it assumes that the conditional distribution of the error given the independent variables has:

- mean 0, $\mathbb{E}[\epsilon_i \mid X] = 0$;
- known variance Ω, $\mathrm{Var}[\epsilon_i \mid X] = 0$.

The GLS Estimator

Recall that the residual vector is: $\epsilon = Y - \hat{Y} = Y - X\beta$

The GLS "selects" the sole coefficient β to minimise the **Mahalanobis distance** of the residual:

$$\min_{\beta}(Y - X\beta)'\Omega(Y - X\beta)$$

We can find a formula for the slope coefficient:

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$$

This estimator has nice statistical properties. It is:

- unbiased,
- consistent,
- efficient, and
- asymptotically normal: $\sqrt{n}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N\left(0,(X'\Omega^{-1}X)^{-1}\right)$.

Connection GLS and OLS

We can derive the GLS estimator the OLS we derived earlier.

To that end,

1. Decompose the covariance matrix as $\Omega = CC'$ using a Cholesky decomposition.
2. Premultiply the linear equation $Y = X\beta + \epsilon$ by C^{-1} to get a new linear model
$$Y^G = X^G\beta + \epsilon^G,$$
where $Y^G = C^{-1}Y,\ X^G = C^{-1}X,\ \epsilon^G = C^{-1}\epsilon.$

3. Observe that
$$\mathrm{Var}\left[\epsilon|X\right] = C^{-1}\Omega C^{-1} = I,$$
the identity matrix (matrix with 1 on the main diagonal and 0 everywhere else.
4. Apply the OLS method to
$$(Y^G - X^G\beta)'(Y^G - X^G\beta) = (Y - X\beta)'\Omega(Y - X\beta)$$
to get an estimate for the slope coefficient as
$$\hat{\beta} = (X'X)^{-1}X'Y = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y.$$

Remark that the transformation in step 2 is actually:

- a rotation, which addresses the autocorrelation of residuals, and
- a rescaling, which addresses the heteroskedasticity of residuals.

Part III - Optimization With Equality Constraints

With linear constraints, the optimization problem becomes:

$$\min_{x_1,\ldots,x_n} f(x_1,\ldots,x_n)$$

subject to:

$$\begin{gathered} g_1(x_1,\ldots,x_n) = b_1 \\ \vdots \\ g_m(x_1,\ldots,x_n) = b_m \end{gathered}$$

Obviously, now we cannot use standard calculus to solve our minimization.

However, wouldn't it be nice if we could transform our objective function f to somehow integrate all the constraints so that we can just apply standard calculus to the new function to solve the problem?

Yes, we can! Thanks to the **Lagrange method**.

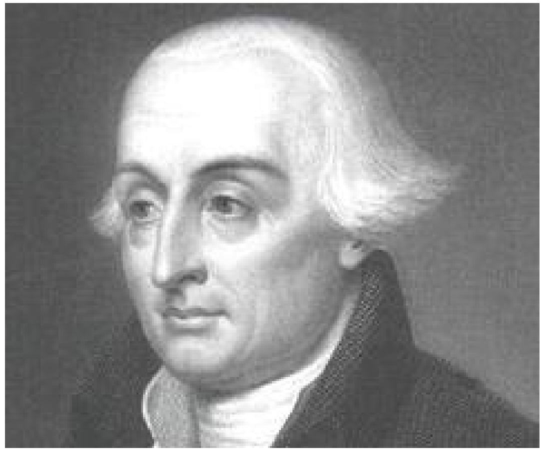


Figure: Joseph-Louis Lagrange (1736-1813), born Giuseppe Luigi Lagrangia

The method of Lagrange

First form the **Lagrangian function** $L(x, \lambda)$.

L is our objective function f augmented by the addition of the constraint functions.

Each constraint function is multiplied by a variable, called a **Lagrange multiplier**.

$$L(x, \lambda) = f(x) + \sum_{j=1}^{m} \lambda_j (g_j(x) - b_j)$$

We use the vectors $x = (x_1, \ldots, x_n)$ and $\lambda = (\lambda_1, \ldots, \lambda_m)$ where appropriate to lighten the notation.

The Lagrange function effectively transforms a problem in n variables $(x_1, \dots, x_n)$ and m constraints into an unconstrained optimization with $n + m$ variables $(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m)$.

We now reformulate our optimization problem in terms of the Lagrange function:

$$\min_{x,\lambda} L(x, \lambda)$$

and use standard calculus to solve it.

By the first order condition, we get a system of $n+m$ equations in $n+m$ unknowns:

$$\frac{\partial L}{\partial x_i}(x) = \frac{\partial f}{\partial x_i}(x) + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i}(x) = 0$$

$$i = 1, \ldots, n$$

$$\frac{\partial L}{\partial \lambda_j}(x) = g_j(x) - b_j = 0 \qquad j = 1, \ldots, m$$

Note that:

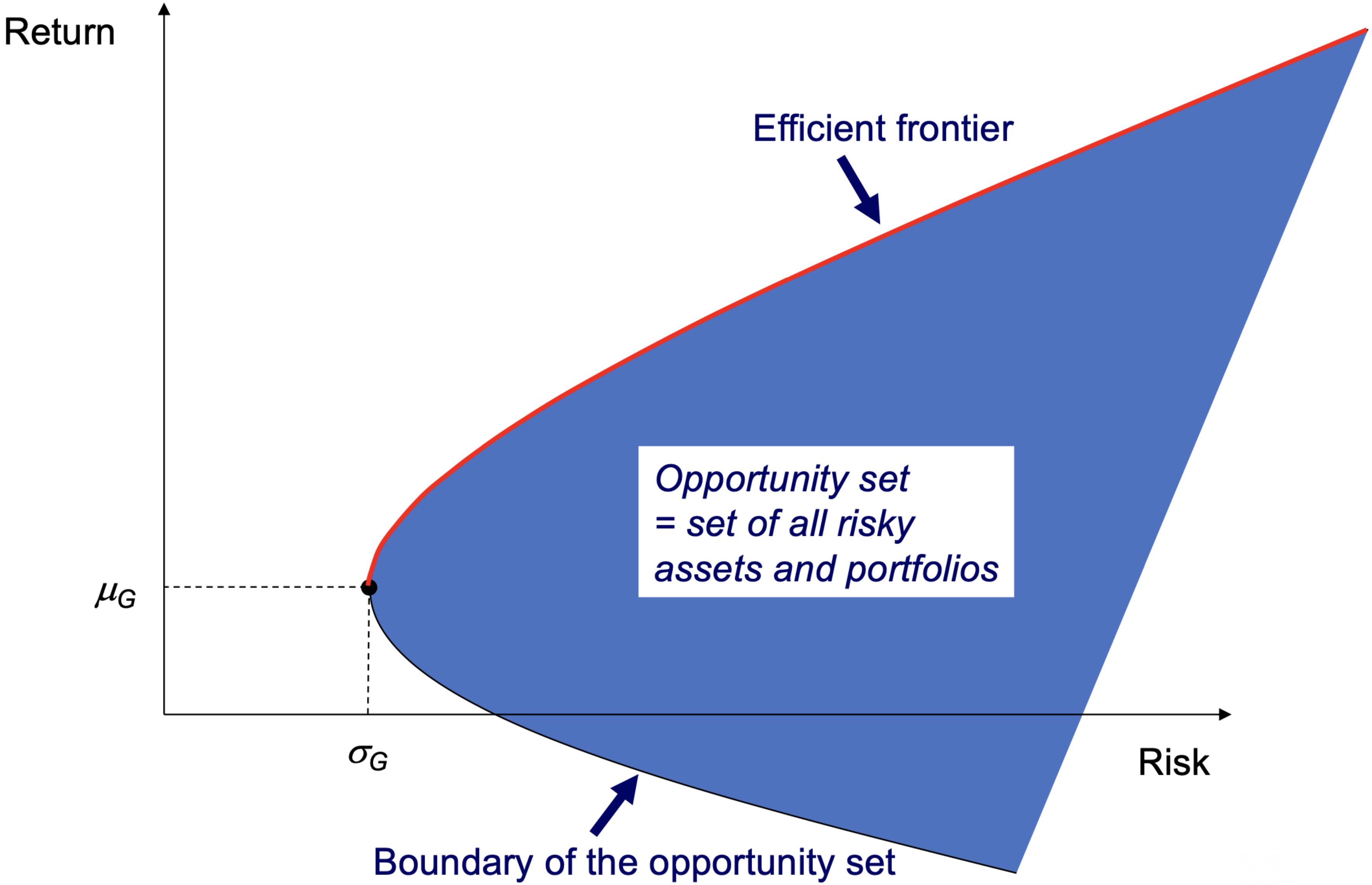
- the first n equations correspond to an unconstrained optimization over f penalized by a sum of functions parameterized by λ;
- we have recovered our constraints in the last m equations.

Once the system of equations is solved, all we need to do is check the Hessian of f before concluding[1].

We will see two portfolio-related applications in the next part.

[1] Strictly speaking, checking the Hessian is not enough in general optimization problems, but it will prove sufficient for the types of applications we consider in this lecture.

Risk Minimization With N Risky Assets



The portfolio selection problem is generally defined as a minimization of risk subject to a return constraint. Two reasons for this convention are:

- a return objective seems intuitively easier to formulate than a risk objective;
- risks are easier to control than returns;

If we adopt this convention, our objective function is the portfolio variance, and we will minimize it with respects to the portfolio weights.

Actually, instead of using the portfolio variance, we will use a little trick and scale it down by a factor of $1/2$ to ease our calculations. Since the factor is positive, it does not affect the value of the optimal vector of weights w^*.

$$\min_w \frac{1}{2}\sigma_\pi^2 = \frac{1}{2}w'\Sigma w$$

Now for our constraints.

We already have one constraint: the portfolio return must be equal to a prespecified level m. In mathematical terms:

$$\mu_\pi = \mu' w = w' \mu = m$$

We also have a second constraint on the weights called the 'budget equation'. The sum of all the weights must necessarily equal 1. Since there is no risk-free assets, our wealth must be entirely invested in a combination of the n assets.

$$\mathbf{1}' w = w' \mathbf{1} = 1$$

where $\mathbf{1}$ is a n-element unit vector:

$$\mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

Summing it all, we formulate the portfolio selection problem as

$$\min_{w} \frac{1}{2} w' \Sigma w$$

Subject to:

$$\begin{aligned} w'\mu &= m \\ w'\mathbf{1} &= 1 \end{aligned}$$

This problem is an optimization with equality constraints. We can solve it using the method of Lagrange.

We form the Lagrange function with two Lagrange multipliers λ and γ:

$$L(w, \lambda, \gamma) = \frac{1}{2} w' \Sigma w + \lambda(m - w'\mu) + \gamma(1 - w'\mathbf{1})$$

Next, we solve for the first order condition by taking the derivative with respect to the vector w:

$$\frac{\partial L}{\partial w}(w, \lambda, \gamma) = \Sigma w - \lambda \mu - \gamma \mathbf{1} = 0 \tag{19}$$

Checking the second order condition, the Hessian of the objective function is equal to the covariance matrix Σ, which is positive definite. Therefore, we have reached the optimal weight vector w^*:

$$w^* = \Sigma^{-1}(\lambda \mu + \gamma \mathbf{1}) \tag{20}$$

To get to this relationship, we have premultiplied (19) by the inverse matrix Σ^{-1}.

All we have to do now is to find the values for λ and γ and then substitute them into (20).

Remember that the constraints are:

$$\begin{aligned} \mu' w &= m \\ \mathbf{1}' w &= 1 \end{aligned}$$

Substituting w^* into these two equations, we get:

$$\mu' \Sigma^{-1}(\lambda\mu + \gamma\mathbf{1}) = \lambda\mu'\Sigma^{-1}\mu + \gamma\mu'\Sigma^{-1}\mathbf{1} = m$$

$$\mathbf{1}'\Sigma^{-1}(\lambda\mu + \gamma\mathbf{1}) = \lambda\mathbf{1}'\Sigma^{-1}\mu + \gamma\mathbf{1}'\Sigma^{-1}\mathbf{1} = 1$$

For convenience, we define the following scalars:

$$\begin{cases} A = \mathbf{1}'\Sigma^{-1}\mathbf{1} \\ B = \mu'\Sigma^{-1}\mathbf{1} = \mathbf{1}'\Sigma^{-1}\mu \\ C = \mu'\Sigma^{-1}\mu \end{cases}$$

Note also that $AC - B^2 > 0$.

The previous system of equations for the Lagrange multipliers becomes

$$\begin{aligned} C\lambda + B\gamma &= m \\ B\lambda + A\gamma &= 1 \end{aligned}$$

Then

$$\begin{cases} \lambda = \frac{Am - B}{AC - B^2} \\ \gamma = \frac{C - Bm}{AC - B^2} \end{cases} \tag{21}$$

Now all we need to do is to substitute these values back into (20) to obtain w^*:

$$w^* = \frac{1}{AC - B^2}\boldsymbol{\Sigma}^{-1}\left[(A\mu - B\mathbf{1})m + (C\mathbf{1} - B\mu)\right] \tag{22}$$

Numerical Application

Consider a market with 4 asset classes X_1, X_2, X_3 and X_4. Their return vector and standard deviation vector are respectively given by

$$\mu = \begin{pmatrix} 0.05 \\ 0.07 \\ 0.15 \\ 0.27 \end{pmatrix}$$

and

$$\sigma = \begin{pmatrix} 0.07 \\ 0.12 \\ 0.30 \\ 0.60 \end{pmatrix}$$

The correlation between asset returns is given by

$$R = \begin{pmatrix} 1 & 0.8 & 0.5 & 0.4 \\ 0.8 & 1 & 0.7 & 0.5 \\ 0.5 & 0.7 & 1 & 0.8 \\ 0.4 & 0.5 & 0.8 & 1 \end{pmatrix}$$

What is the optimal asset allocation to obtain a return $m = 10\%$

Answer:

The first step is to create the covariance matrix Σ:

$$\Sigma = SRS = \begin{pmatrix} 0.0049 & 0.00672 & 0.0105 & 0.0168 \\ 0.00672 & 0.0144 & 0.0252 & 0.036 \\ 0.0105 & 0.0252 & 0.09 & 0.144 \\ 0.0168 & 0.036 & 0.144 & 0.36 \end{pmatrix}$$

Next, we compute A, B and C...

$$\begin{cases} A = \mathbf{1}'\Sigma^{-1}\mathbf{1} = 239.3440468 \\ B = \mu'\Sigma^{-1}\mathbf{1} = 9.618456078 \\ C = \mu'\Sigma^{-1}\mu = 0.550280982 \end{cases}$$

... as well as λ well as γ

$$\begin{cases} \lambda = 0.3652794 \\ \gamma = -0.0105013 \end{cases}$$

Now, we can answer our question. The optimal asset allocation to obtain a return $m = 10\%$ is given by

$$\begin{aligned} w^* &= \Sigma^{-1}(\lambda\mu + \gamma\mathbf{1}) \\ &\approx \Sigma^{-1}(0.3652794\mu + 0.0105013) \\ &\approx \begin{pmatrix} 0.528412108 \\ 0.172888075 \\ 0.159764343 \\ 0.138935474 \end{pmatrix} \end{aligned}$$

If we were to double check with Excel Solver, we would find that

$$w^* \approx \begin{pmatrix} 0.528412169 \\ 0.172888796 \\ 0.159764425 \\ 0.138935450 \end{pmatrix}$$

As expected, the theory works!

The Minimum-Variance Portfolio

To find the minimum-variance portfolio, we could directly solve the following portfolio optimization problem:

$$\min_{w} \frac{1}{2} w' \Sigma w$$

Subject to:

$$w' \mathbf{1} \quad = \quad 1.$$

While this direct approach works well, it ignores the results that we have obtained so far.

Here, we present an alternative way of obtaining the minimum-variance portfolio that uses the solution of our previous, and general, problem with a return target.

In our previous problem, the weights of the optimal portfolio depend on the return objective m specified in the constraints.

By letting m span the real line, we have obtained the minimum-variance frontier, aka the boundary of the opportunity set. To identify the global minimum-variance portfolio on this frontier, we just need to find its expected return. Then the asset allocation and variance will follow.

That's what we do here.

We start by expressing the variance of a portfolio located on the boundary of the opportunity set as a function of m:

$$\sigma_\pi^2(m) = \frac{Am^2 - 2Bm + C}{AC - B^2} \tag{23}$$

How did we get there?

Getting to this expression for $\sigma_\pi^2(m)$ is simple but tedious.

Start from the equation for the variance of a portfolio:

$$\sigma_\pi^2 = w'\Sigma w \tag{24}$$

Notice that the asset allocation of the portfolios located on the boundary of the opportunity set (and therefore on the efficient frontier!) is fully parametrised by the target return of the portfolio. So we can express equation (20) as

$$w^*(m) = \Sigma^{-1}(\lambda(m) \cdot \mu + \gamma(m) \cdot \mathbf{1}), \tag{25}$$

where

$$\begin{cases} \lambda(m) = \frac{Am - B}{AC - B^2} \\ \gamma(m) = \frac{C - Bm}{AC - B^2} \end{cases} \tag{26}$$

(recall that A, B, C are just scalar parameters!)

Now, substitute (26) into (25) and plug the result into (24) to get

$$\sigma_\pi^2(m) = \frac{Am^2 - 2Bm + C}{AC - B^2} \tag{27}$$

The **minimum variance portfolio** is the portfolio solving the unconstrained problem:

$$\min_{m} \sigma_{\pi}^{2}(m) \tag{28}$$

Solving for the first-order condition, we get the candidate solution:

$$\frac{d\sigma_{\pi}^{2}(m)}{dm} = 0 \Leftrightarrow \frac{2Am - 2B}{AC - B^2} \Leftrightarrow m = \frac{B}{A}$$

Checking the second-order condition confirms that the candidate is the actual solution m_g:

$$m_g = \frac{B}{A}$$

Now that we have m_g, we can easily get the vector of weights w_g:

$$w_g = \frac{\Sigma^{-1}\mathbf{1}}{A}$$

and finally, the variance:

$$\sigma_g^2 = \frac{1}{A}$$

Numerical Example... continued...

What is the global minimum variance portfolio's asset allocation? What are its return and standard deviation?

Answer:

The global minimum variance portfolio's asset allocation is given by

$$w_g = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{A} \approx \begin{pmatrix} 1.274886723 \\ -0.263112728 \\ 0.016339421 \\ -0.028113415 \end{pmatrix}$$

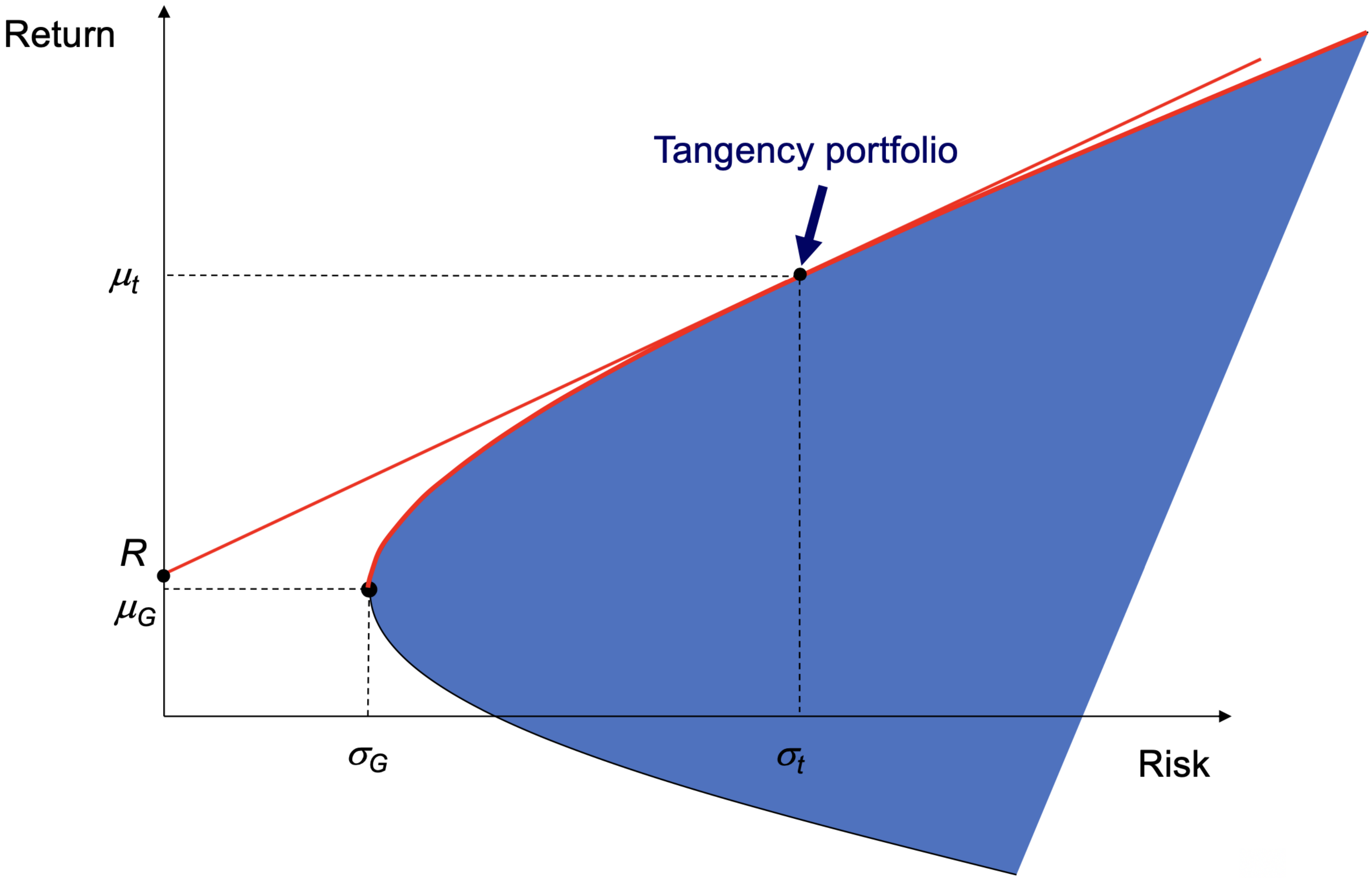
its return is

$$m_g = \frac{B}{A} \approx 4.02\%$$

and its standard deviation is equal to

$$\sigma_g = \sqrt{w_g' \Sigma w_g} \approx 6.46\%$$

Risk Minimization With N Risky Assets and a Risk-Free Asset



We move to the next optimization problem on our list: how should we allocate our wealth to N risky assets and a risk-free asset when we have a return target, and we want to minimize the portfolio's risk subject to that target.

When we have N risky assets and a risk-free asset, the optimization problem is simpler than when we do not have a risk-free asset because the budget constraint vanishes because the residual of the wealth not invested in risky assets is now in the risk-free asset. Moreover, the return equation is now (2).

So the optimization problem becomes:

$$\min_{w} \frac{1}{2} w' \Sigma w$$

Subject to:

$$r + w'(\mu - r\mathbf{1}) = m$$

We form the Lagrange function

$$L(x, \lambda) = \frac{1}{2} w' \Sigma w + \lambda (m - r - w'(\mu - r\mathbf{1})),$$

with Lagrange multiplier λ to account for the equality constraint.

Next, we solve for the first-order condition by taking the derivative of L with respect to the vector w:

$$\frac{\partial L}{\partial w} = \Sigma w - \lambda(\mu - r\mathbf{1}) = 0$$

Checking the second-order condition, the Hessian of the objective function is still the covariance matrix, which is positive definite. Therefore, we have reached the optimal weight vector w^*:

$$w^* = \lambda \Sigma^{-1}(\mu - r\mathbf{1}) \tag{29}$$

Here again, we have transposed our equation and multiplied both sides by the inverse matrix Σ^{-1}.

Substituting the value of w^* at (29) into the constraint,

$$\begin{aligned} & r + w'(\mu - r\mathbf{1}) = m \\ \Leftrightarrow & r + \lambda(\mu - r\mathbf{1})'\Sigma^{-1}(\mu - r\mathbf{1}) = m \end{aligned} \tag{30}$$

Next, we solve for λ to get the value of the Lagrange multiplier:

$$\lambda = \frac{m - r}{(\mu - r\mathbf{1})'\Sigma^{-1}(\mu - r\mathbf{1})}$$

Substituting the value we found for λ into equation (29), we get the final expression for w^* as:

$$w^* = \frac{(m - r)\Sigma^{-1}(\mu - r\mathbf{1})}{(\mu - r\mathbf{1})'\Sigma^{-1}(\mu - r\mathbf{1})} \tag{31}$$

Numerical Example ...continued...

Assume the risk-free rate is 2.5%. What is the allocation of the portfolio returning 10%?

Answer:

The allocation of the portfolio returning 10%? is given by

$$w^* = \frac{(m - r)\Sigma^{-1}(\mu - r\mathbf{1})}{(\mu - r\mathbf{1})'\Sigma^{-1}(\mu - r\mathbf{1})} \approx \begin{pmatrix} 0.887352483 \\ 0.08126325 \\ 0.15484343 \\ 0.121648624 \end{pmatrix}$$

The Tangency Portfolio

The **tangency portfolio** is the portfolio that is entirely invested in risky assets.

The vector of weights for the tangency portfolio, w_t is given by:

$$w_t = \frac{\Sigma^{-1}(\mu - r\mathbf{1})}{B - Ar}$$

The mean and standard deviation of this portfolio are given by

$$m_t = w_t'\mu = \frac{C - Br}{B - Ar}$$

and

$$\sigma_t = \sqrt{w_t'\Sigma w_t} = \sqrt{\frac{C - 2rB + r^2 A}{(B - Ar)^2}}$$

- How did we get there?

To derive the asset allocation for the tangency portfolio, notice first that the tangency portfolio is very special because it is simultaneously:

- on the **new** efficient frontier for risky and risk-free assets (the Capital Market Line). Hence, its asset allocation satisfies equation (31):

$$w_t = \frac{(m_t - r)\Sigma^{-1}(\mu - r\mathbf{1})}{(\mu - r\mathbf{1})'\Sigma^{-1}(\mu - r\mathbf{1})}$$

- on the **old** risky-only efficient frontier (the hyperbola!). So, its asset allocation also satisfies equation (20), that is:

$$w_t = \Sigma^{-1}(\lambda\mu + \gamma\mathbf{1})$$

  where λ and γ are given at (21).

Because the tangency portfolio is fully invested in risky assets, then its asset allocation must satisfy the budget equation:

$$\mathbf{1}' w_t = 1$$

Substituting (31), we get

$$\mathbf{1}' \frac{(m_t - r)\Sigma^{-1}(\mu - r\mathbf{1})}{(\mu - r\mathbf{1})'\Sigma^{-1}(\mu - r\mathbf{1})} = 1$$

Using the definitions of A, B and C in the denominator, this condition becomes

$$\mathbf{1}' \frac{(m_t - r)\Sigma^{-1}(\mu - r\mathbf{1})}{C - 2rB + r^2 A} = 1$$

Multiplying both sides by $C - 2rB + r^2A$, we get:

$$(m_t - r)\mathbf{1}'\mathbf{\Sigma}^{-1}(\mu - r\mathbf{1}) = C - 2rB + r^2A$$

Using again the definitions of A, B and C and simplifying, we have:

$$m_t(B - rA) = C - rB$$

This gives us the formula for the return of the tangency portfolio:

$$m_t = \frac{C - Br}{B - Ar}$$

Now, to get w_t, simply plug the formula for m_t in equation (31) and use the definition of A, B and C to simplify:

$$w_t = \frac{(m_t - r)\Sigma^{-1}(\mu - r\mathbf{1})}{(\mu - r\mathbf{1})'\Sigma^{-1}(\mu - r\mathbf{1})} = \frac{\Sigma^{-1}(\mu - r\mathbf{1})}{B - Ar}$$

Numerical Example ...concluded

What is the tangency portfolio's asset allocation? What are its return and standard deviation?

Answer:

The asset allocation of the tangency portfolio is given by

$$w_t = \frac{\Sigma^{-1}(\mu - r\mathbf{1})}{B - Ar} \approx \begin{pmatrix} 0.712671217 \\ 0.065266037 \\ 0.124361466 \\ 0.097701279 \end{pmatrix}$$

its return is

$$m_t = \frac{C - Br}{B - Ar} \approx 0.085235749$$

and its standard deviation is equal to

$$\sigma_t = \sqrt{\frac{C - 2Br + Ar^2}{(B - Ar)^2}} \approx 0.128731141$$

Part III: The Black-Litterman Model

Imagine you are working in the asset management division of a large bank. Your daily challenge is to come up with portfolios that will provide superior risk-adjusted return to your investors.

A few floors down, in the investment banking divisions, dozens of equity and fixed income analysts pour over thousands of financial statements, economic releases and geopolitical news to provide buy and sell recommendations to their clients. How could you harness the views formulated by these analysts to improve your portfolio construction?

This is the challenge that Fisher Black and Robert Litterman successfully tackled in the early 1990s. Their model, the Black-Litterman model published in 1992 [1] is still the reference.

Other resources include articles by He and Litterman [5], Walters [8] (of *blacklitterman.org* fame), Thomas Idzorek [4], and Alles Rodrigues and Lleo [6]. The Black-Litterman approach has also been adapted to dynamic portfolio selection models.

Figure: Fisher Black (1938-1995)

Figure: Robert Litterman

Formalizing the Black-Litterman Model

Throughout, we will be looking at excess returns, as opposed to nominal returns.

We start from a simple assumption: the excess returns $\tilde{R} = R - r\mathbf{1}$ are Normally distributed with mean $\tilde{\mu} = \mu - r\mathbf{1}$ and covariance Σ, that is:

$$\tilde{R} \sim N(\tilde{\mu}, \Sigma)$$

The basic ideas in Black-Litterman is that the 'true' mean excess return $\tilde{\mu}$ is neither known with certainty nor directly observable. So we have to relay on a noisy estimate π. We model its distribution as:

$$\pi \sim N(\tilde{\mu}, \Sigma_\pi)$$

Stated otherwise, π is a noisy (random) estimate gravitating around the 'true mean' $\tilde{\mu}$, that is

$$\pi = \tilde{\mu} + \epsilon \tag{32}$$

where $\epsilon \sim N(0, \Sigma_\pi)$.

If we assume that ϵ and $\tilde{\mu}$ are uncorrelated we deduce that

$$\tilde{R} \sim N(\tilde{\mu}, \Sigma_r)$$

where the variance is $\Sigma_r = \Sigma + \Sigma_\pi$.

Black-Litterman: Step by Step

We can derive the Black-Litterman formula using either of two methods:

- **Theil's mixed estimation model** (see [7]). This is the technique initially used by Black and Litterman in their landmark article [1]. Theil's approach relies on a generalised least squares regression (GLS), which is an extension of the ordinary least square (OLS) method we derived at the beginning of this class.
- **Bayes's Formula**: the idea behind Black-Litterman, updating an uninformed base asset allocation using new information contained in the views of your analysts to get a new an improved asset allocation, can be naturally explained in the language of Bayesian statistics.

In the following slides, we will use the Bayesian approach to construct the Black-Litterman formula step by step.

A Black-Litterman (Bayesian) roadmap:

- Step 0: Introducing Bayes' Formula;
- Step 1: Reverse optimization to get the prior;
- Step 2: Inputing the view;
- Step 3: Combining to get the posterior distribution;
- Step 4: Asset allocation.

Step 0: An introduction to Bayes' Formula

- Bayes' formula is closely related to the multiplication rule and the total probability rule (see Appendix).
- Bayes' formula is routinely used to update probabilities when new information becomes available:

$$\begin{aligned} & \text{Updated probability} \\ = \; & \frac{\text{Probability of a new information for a given event}}{\text{Unconditional probability of new information}} \\ & \times \text{Prior probability of the event} \end{aligned}$$

Mathematically,

$$P(E|I) = \frac{P(I|E)}{P(I)} \times P(E)$$

where E is an event and I is the new piece of information.

Baye's Formula is a direct application of the multiplication rule:

$$P(I) \times P(E|I) = P(E) \times P(I|E)$$

A bit of jargon:

- $P(E)$ is called the **prior** probability. It is the uninformed probability of event E.
- $P(E|I)$ is called the **posterior** probability and represents the probability of event E once we have incorporated information I.
- $P(I)$ is a **normalisation constant**.

The 1950s witnessed a renewed interest in Bayes's formula, due in no small part to Leonard Jimmie Savage.

- Mathematician and statistician
- Worked closely with John Von Neumann during World War II
- Friend to both Paul Samuelson and Milton Friedman (!)
- Gave Samuelson the reference to Bachelier's thesis... on the back of a postcard
- Popularized the use of Bayes' formula through his book 'Foundations of Statistics' published in 1954.

Figure: Leonard Jimmie Savage (1917-1971)

Step 1: Reverse optimization to get the prior

For starters, we need a **prior** distribution for the excess returns. Here, the prior is a base case asset allocation. It should represent a neutral and uninformed view of the market.

The trouble is that there is no unique choice: one could start from

- a uniform ("1/N") portfolio;
- the global minimum variance portfolio;
- any other 'neutral and uninformed' portfolio.

However, these starting points are far from ideal:

- Uniform portfolios cannot really help us establish a prior view of excess return;
- To get the returns of global minimum variance, we implicitly need an estimate for the excess returns. If we use historical returns, we incur the risk of having estimates that do not reflect future expectations, and that might result in unrealistic portfolios;

But Black and Litterman's solution to the prior selection quandary is both effective and elegant: they start from the equilibrium CAPM portfolio.

To get the prior expected excess return, we will start from the mean-variance problem

$$\max_{w} w'\tilde{R} - \frac{\lambda}{2} w'\Sigma w$$

where Σ represents the covariance matrix of excess returns.

The solution to this problem is

$$w^* = \frac{1}{\lambda}\Sigma^{-1}\tilde{R} \tag{33}$$

Now, the trouble is that we do not really know what the 'real' (equilibrium) vector of excess return is....

However, **<u>if</u>** we assume that the CAPM holds, meaning that markets are (broadly) in equilibrium, then we can reverse engineer the equilibrium vector of excess return Π out of equation (34).

Specifically,

$$w_{\text{mkt}} = \frac{1}{\lambda_{\text{mkt}}}\Sigma^{-1}\Pi \tag{34}$$

where w_{mkt} are the weights of the market portfolio reached in equilibrium, λ_{mkt} is the (equilibirum) risk-aversion of market participants, and Π is the equilibrium vector of risk premia. Then,

$$\Pi = \lambda_{\text{mkt}}\Sigma w_{\text{mkt}} \tag{35}$$

Concretely, we would use the current asset allocation of our favourite broad market-cap weighted index. The argument is that if we are close to equilibrium, any broad market index should reflect the market portfolio.

How to?... get the risk aversion λ

We may not know what the risk aversion λ_{mkt} is, but we could estimate it. Start by premultiplying (35) by w'_{mkt}:

$$w'_{\text{mkt}}\Pi = \lambda_{\text{mkt}} w'_{\text{mkt}} \Sigma w_{\text{mkt}} \tag{36}$$

Now, $w'_{\text{mkt}}\Sigma w_{\text{mkt}} =: \sigma^2_{\text{mkt}}$ is the variance of the market portfolio and $w'_{\text{mkt}}\Pi$ is the excess return of the market portfolio.

Thus,

$$\lambda_{\text{mkt}} = \frac{w'_{\text{mkt}}\Pi}{\sigma^2_{\text{mkt}}} = \frac{1}{\sigma_{\text{mkt}}} \mathcal{S}_{\text{mkt}} \tag{37}$$

where $\mathcal{S}_{\text{mkt}} = \frac{w'_{\text{mkt}}\Pi}{\sigma_{\text{mkt}}}$ is the Sharpe ratio of the market.

Black and Litterman use a Sharpe ratio close to 0.5 in their article.

This reverse optimization is actually a very nice idea, with several key advantages:

- the prior portfolio allocation is neutral and uninformed: it is just the allocation of a broad market index!
- the prior portfolio allocation is realistic: it does not contain extreme weights;
- we obtain a prior excess return: the equilibrium vector of excess returns Π.

Moreover, formula (34) has shown us that there is a direct (and simple) relation linking excess returns and portfolio weights.

Numerical Example

We return to our market with 4 assets classes X_1, X_2, X_3 and X_4. The (historical) covariance matrix of the asset classes Σ:

$$\Sigma = \begin{pmatrix} 0.0049 & 0.00672 & 0.0105 & 0.0168 \\ 0.00672 & 0.0144 & 0.0252 & 0.036 \\ 0.0105 & 0.0252 & 0.09 & 0.144 \\ 0.0168 & 0.036 & 0.144 & 0.36 \end{pmatrix}$$

We have obtained this covariance matrix using 120 monthly returns (10 years).

We also identify a broad (market cap weighted) index with allocation

$$w_{\text{mkt}} = \begin{pmatrix} 0.05 \\ 0.40 \\ 0.45 \\ 0.10 \end{pmatrix}$$

We can compute the equilibrium vector of excess return Π (the prior) using equation (35). First, we need to estimate λ_{mkt} using equation (37).

The standard deviation of the market is

$$\sigma_{\mathrm{mkt}} = \sqrt{w'_{\mathrm{mkt}} \Sigma w_{\mathrm{mkt}}} = 22.35\%$$

Taking a Sharpe ratio of 0.5 (the same as Black and Litterman), we obtain

$$\lambda = \frac{1}{\sigma_{\mathrm{mkt}}} \mathcal{S}_{\mathrm{mkt}} = \frac{1}{0.2235} \times 0.5 = 2.24$$

The equilibrium vector of excess return Π (the prior) subject to a risk aversion $\lambda_{\text{mkt}} = 2.24$ is

$$\Pi = \lambda_{\text{mkt}} \Sigma w_{\text{mkt}} = \begin{pmatrix} 0.0209 \\ 0.0471 \\ 0.1465 \\ 0.2596 \end{pmatrix}$$

... and now, back to the theory...

At this stage, we have (almost) obtained a **prior distribution** for the excess return

$$P(E) \sim N(\Pi, \Sigma_\pi)$$

The difficulty is that we do not know what Σ_π is...

To get things going, Black and Litterman assume that the variance of the estimate Σ_π is proportional to the covariance matrix of the excess returns Σ with a coefficient of proportionality τ:

$$\Sigma_\pi = \tau\Sigma$$

With this assumption in mind, the **prior distribution** for the excess return is

$$P(E) \sim N(\Pi, \tau\Sigma) \tag{38}$$

How to?... specify the parameter τ

There is no clear guidance on how to chose τ. Generally, τ will be close to 0, with some authors advocating values between 0.01 and 0.05.

If we estimate our model using a time series of T historical data, we could view $\tau\Sigma$ as the square of the *standard error of estimate* for the equilibrium vector Π. We would therefore pick $\tau \approx T^{-1}$.

Numerical Example

To get the variance of the prior, we compute τ based on the standard error of estimate method:

$$\tau = \frac{1}{T} = \frac{1}{120} = 0.0083 \tag{39}$$

slightly below the 0.01 - 0.05 range.

The prior distribution is

$$P(E) \sim N(\Pi, \tau\Sigma) \tag{40}$$

with

$$\Pi = \begin{pmatrix} 0.0209 \\ 0.0471 \\ 0.1465 \\ 0.2596 \end{pmatrix}$$

and

$$\tau\Sigma = \begin{pmatrix} 0.00004 & 0.00006 & 0.00009 & 0.00014 \\ 0.00006 & 0.00012 & 0.00021 & 0.00030 \\ 0.00009 & 0.00021 & 0.00075 & 0.00120 \\ 0.00014 & 0.00030 & 0.00120 & 0.00300 \end{pmatrix}$$

Step 2: Inputing the views

The main feature of the Black-Litterman model is that it incorporate the views of analysts and portfolio managers to fine tune our estimates of expected returns and our asset allocation.

Here again, Black and Litterman propose an effective and elegant solution that allows both absolute and relative views.

To illustrate the procedure, we use a simple example from Walters [8] with four assets and two views:

1. *Relative view:* the investor believes that Asset 3 will outperform Asset 1 by 10% with confidence ω_1;
2. *Absolute view:* the investor believes that Asset 2 will return 3% with confidence ω_2,

We will express the views as a matrix P of trades or positions:

$$P = \begin{pmatrix} -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Absolute views are expressed as long positions, while relative views are represented long-short positions.

Each view implies a specific return for one or several asset classes. These returns are contained in the vector Q:

$$Q = \begin{pmatrix} 0.10 \\ 0.03 \end{pmatrix}$$

We also need a matrix Ω for the confidence. This matrix will be used as a covariance matrix to model the uncertainty in the views:

$$\Omega = \begin{pmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{pmatrix}$$

More generally, if we have K views on N assets, we will need

- a $K \times N$ matrix of views P. For relative views, the sum of the weights needs to equal 0 while for absolute the sum of the weights will equal 1.
- a K-element column vector of returns Q;
- a $K \times K$ covariance matrix expressing the degree of confidence in the views. We assume that the degree of accuracy of the views are independent, so Ω is a diagonal matrix.

If we assume that the uncertainty about the views is Gaussian, we can express the conditional distribution for the views as:

$$P(I|E) \sim N(Q, \Omega) \tag{41}$$

How to?... specify the covariance matrix Ω

Walter [8] presents four methods that can be used to specify Ω:

1. *Proportional to the variance of the prior*: He and Litterman [3] set Ω as

$$\Omega := Diag(P(\tau\Sigma)P')$$

   where $Diag(.)$ is the diagonalization function.
2. *Use a confidence interval*;
3. *Use the variance of residuals in a factor model*: this is the method of choice if you are using a factor model *à la* Grinold and Kahn (see [2]);
4. *Idzorek's method*: Idzorek [4] proposes an approach to specify the confidence.

Numerical Example

To get the matrix Ω, we use the approach proposed by He and Litterman.

First, compute

$$P(\tau\Sigma)P' = \begin{pmatrix} 0.000615833 & 0.000154 \\ 0.000154 & 0.00012 \end{pmatrix}$$

Next, we diagonalize:

$$\Omega := Diag(P(\tau\Sigma)P') = \begin{pmatrix} 0.000615833 & 0 \\ 0 & 0.00012 \end{pmatrix}$$

Step 3: Combining to get the posterior distribution

So far, we have derived:

- the prior probability $P(E)$;
- the conditional distribution of the views given our prior, $P(I|E)$.

It turns out that we do not need to derive the normalizing constant $P(I)$ explicitly as it will vanish into the integration constant when we perform the calculations. This is generally the case when we apply Bayes' approach to continuous distributions.

To conclude, we 'just' need to plug into Bayes' formula, do a few pages of tedious calculations... to eventually derive the posterior distribution of excess return:

$$P(E|I) \sim N\left(\left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}\left[(\tau\Sigma)^{-1}\Pi + P'\Omega^{-1}Q\right], \left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}\right) \tag{42}$$

The most important information in formula (42) is the posterior expected excess return:

$$\hat{R}_I := \mathbb{E}\left[\tilde{R}|I\right] = \left[(\tau\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}\left[(\tau\Sigma)^{-1}\Pi + P'\Omega^{-1}Q\right] \tag{43}$$

This result is the Black-Litterman formula!

Application

Plugging our results into formula (42), we get:

- the posterior expected excess return:

$$\begin{aligned}\hat{R}_I &:= \mathbb{E}\left[\tilde{R}|I\right] \\ &= \begin{pmatrix} 0.0168 \\ 0.0375 \\ 0.1248 \\ 0.2270 \end{pmatrix}\end{aligned}$$

- the variance of the estimate:

$$\begin{pmatrix} 0.0000276650 & 0.0000272705 & 0.00003 & 0.00006 \\ 0.0000272705 & 0.000054766 & 0.00007 & 0.00009 \\ 0.00003 & 0.00007 & 0.00032 & 0.00053 \\ 0.00006 & 0.00009 & 0.00053 & 0.00196 \end{pmatrix}$$

Step 4: Asset allocation

The final step is to use the posterior expected excess return to derive the optimal asset allocation (conditional on the views expressed).

To do this, we revisit the utility maximisation problem, but this time with our posterior expected excess return $\hat{R}_I$:

$$\max_{w} \left[r + w' \hat{R}_I \right] - \frac{\lambda}{2} w' \Sigma w$$

The resulting asset allocation is

$$w^* = \frac{1}{\lambda} \Sigma^{-1} \hat{R}_I \tag{44}$$

Numerical Example

The ultimate objective of Black-Litterman is to generate new and improved asset allocations.

We compute three asset allocations for three different levels of risk aversion:

- $\lambda = 0.1$, corresponding to an investor with a very low-risk aversion (overbetting);
- $\lambda = 1$, the lowest anyone should go in terms of risk aversion;
- $\lambda = 2.24$, which corresponds to the risk-aversion of market participants that we estimated when we computed the prior;
- $\lambda = 6$, corresponding to a risk-averse investor.

- $\lambda = 0.1$:

$$w^* = \begin{pmatrix} 219.95\% \\ 371.92\% \\ 898.50\% \\ 223.69\% \end{pmatrix}$$

resulting in a **very** large amount (1614%!!!) borrowed at the risk-free rate.

- $\lambda = 1$:

$$w^* = \begin{pmatrix} 22.00\% \\ 37.19\% \\ 89.85\% \\ 22.37\% \end{pmatrix}$$

resulting in a large amount (71.41%) borrowed at the risk-free rate. Still not for the faint-hearted!

- $\lambda = 2.24$:

$$w^* = \begin{pmatrix} 9.83\% \\ 16.63\% \\ 40.17\% \\ 10.00\% \end{pmatrix}$$

with 23.27% invested in the risk-free rate.

- $\lambda = 6$:

$$w^* = \begin{pmatrix} 3.67\% \\ 6.20\% \\ 14.98\% \\ 3.73\% \end{pmatrix}$$

with 71.43%(!) invested in the risk-free rate

Part V: A Short Note on Optimization Problems with Inequality Constraints

We can always write an optimization problem with inequality constraints as:

$$\min_{x_1,\ldots,x_n} f(x_1,\ldots,x_n)$$

subject to:

$$\begin{gathered} g_1(x_1,\ldots,x_n) \leq b_1 \\ \vdots \\ g_m(x_1,\ldots,x_n) \leq b_m \end{gathered}$$

The Kuhn-Tucker theorem provides a set of *necessary* conditions for the existence of an optimal solution.

One possible formulation of the Kuhn-Tucker conditions is:

$$\begin{aligned}
\lambda_0 \frac{\partial f(x)}{\partial x_i} + \sum_{j=1}^{m} \lambda_j \frac{\partial (g_j(x) - b_j)}{\partial x_i} &= 0 \qquad i = 1, \ldots, n \\
\lambda_j (g_j(x) - b_j) &= 0 \qquad j = 1, \ldots, m \\
\lambda_j &\geq 0 \qquad j = 0, \ldots, m
\end{aligned}$$

Underneath the first condition, we recognize the familiar form of the Lagrange function, except that there is an extra Lagrange multiplier. The second set of conditions relates to the constraints, and the third set of conditions relates to the Lagrange multipliers themselves.

Unfortunately, Kuhn-Tucker does not guarantee that a vector x satisfying the condition will be optimal in the general case. The Kuhn-Tucker condition acts as necessary and sufficient condition only when the objective function is convex and the constraints are linear.

Application to Asset Management: 130-30 Portfolio

We can always write an optimization problem with inequality constraints as:

$$\min_{x_1, \ldots, x_n} f(x_1, \ldots, x_n)$$

subject to:

$$\begin{gathered} g_1(x_1, \ldots, x_n) \leq b_1 \\ \vdots \\ g_m(x_1, \ldots, x_n) \leq b_m \end{gathered}$$

Part VI: Extending Mean-Variance Portfolio Optimziation to Benchmarks and Active Portfolio Management

All the problems we have considered so far follow the simple script: "Investor has a pot of money to put in any of N securities. Investors choose the allocation to maximize risk-adjusted returns."
In reality, retail and institutional investors face different situations that do not follow a simple script. For example,

- ETFs need to replicate a benchmark,
- mutual funds attempt to beat their benchmark on a risk-adjusted basis,
- Hedge funds aim at maximizing active returns subject to an active risk constraint,
- Insurance companies need to manage their asset portfolio to guarantee that they will be able to meet their liabilities.

Extensions of mean-variance optimization address these and other practical problems.

A popular way of dealing with a benchmark is to consider your portfolio as the combination of two strategies: a purely passive investment in the benchmark with return R_B and an active zero net weight long-short portfolio with return R_A. Hence, the return of your portfolio $R_p = w_P' R$ is also

$$R_p = R_B + R_A = w_B R + \Delta w' R,$$

where

- w_B is the weight vector for the benchmark, and;
- $\Delta w := w_P - w_B$ are the **active weights**.

The zero net weight constraint on the active portfolio implies that $\Delta w \mathbf{1}' = 1$. Thus, portfolio optimization only concerns the active strategy.

$$\max_{\Delta w} \Delta w' \mu - \frac{\lambda}{2} \Delta w' \Sigma \Delta w$$

Subject to:

$$\Delta w \mathbf{1}' = 1$$

In this lecture, we have seen...

I. Introduction to optimization:
- how to formulate an optimization problem;
- elementary rules and tips.

II. Unconstrained optimization problems
- how to use calculus to solve unconstrained optimization problems;
- application to linear regression;
- application to mean-variance optimization.

III. Optimization problems with equality constraints
- the method of Lagrange;
- application to portfolio selection;
- the minimum variance portfolio;
- the tangency portfolio.

IV. The Black-Litterman model.
V. A short note on optimization problems with inequality constraints: the Kuhn-Tucker conditions.
VI. Extending Mean-Variance Portfolio Optimziation to benchmarks and active portfolio management.

Appendix: Conditional Probability, the Multiplication Rule and the Total Probability Rule

The *conditional probability* of an event E is the probability that E occurs given that another event, say F, has occurred. The conditional probability of E given F is denoted by $P(E|F)$.

The **multiplication rule** states that the probability that A and B both occur equals the probability that A occurs multiplied by the conditional probability that B occurs given that A has occurred:

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

We often use this relation to deduce the conditional probabilities:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad P(B|A) = \frac{P(A \cap B)}{P(A)}$$

When A and B are two **independent** events meaning that the occurrence of either does not affect the likelihood of the other, then

$$P(A|B) = P(A) \qquad P(B|A) = P(B)$$

and as a result

$$P(A \cap B) = P(A)P(B)$$

We can extend the multiplication rule to a set of mutually exclusive and exhaustive events $B_1, B_2, \ldots, B_M$:

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_M)P(B_M)$$

this is called the **total probability rule**.

Where does the total probability rule come from?

Because $B_1, B_2, \ldots, B_M$ are mutually exclusive and exhaustive

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_M)$$

Applying the multiplication rule, we get

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_M)P(B_M)$$

F. Black and R. Litterman.
Global portfolio optimization.
Financial Analyst Journal, 48(5):28–43, Sep/Oct 1992.

R. Grinold and R. Kahn.
Active Portfolio Management : A quantative approach for producing superior returns and selecting superior money managers.
McGraw-Hill, 2 edition, 1999.

G. He and R. Litterman.
The intuition behind black-litterman model portfolios.
Goldman Sachs Asset Management Working paper, 1999.

T. Idzorek.
A step-by-step guide to the black-litterman model, incorporating user-specified confidence levels.
Technical report, Working Paper, 2004.

R. Litterman and Goldman Sachs Asset Management the Quantitative Research Group.
Modern Investment Management: An Equilibrium Approach.
Wiley, 2003.

Alexandre Alles Rodrigues and Sébastien Lleo.
Combining standard and behavioral portfolio theories: a practical and intuitive approach.
Quantitative Finance, 18(5):707–717, 2018.

H. Theil.
Principles of Econometrics.
Wiley, 1971.

J. Walters.
The black-litterman model in detail.
Technical report, Working Paper, 2011.