PORTFOLIO CHOICE AND ASSET PRICING

Q.E.M., M.M.E.F., M.A.E.F., I.M.-M.A.E.F.

Isabelle Nagot
Université PARIS I Panthéon-Sorbonne

Bibliography:

· Elton, E.J., Gruber, M.J., Brown, S.J., and W.N. Goetzmann
Modern Portfolio Theory and Investment Analysis, 8th ed. Wiley, 2011.

· Copeland, T.E., and J.F. Weston *Financial Theory and Corporate Policy*, Addison-Wesley, 1988.

· Back, K.E., *Asset Pricing and Portfolio Choice Theory*, Oxford University Press, 2010.

· P. Viala et E. Briys *Eléments de théorie financière.* Nathan, 1995.

· Bodie Z., A. Kane and A. Marcus *Investments*, 8th Edition, 2009.

Abstract Basic principles underlying rational portfolio choice
+ equilibrium conditions in the capital markets to which the previous analysis leads.

Chapter I Portfolio analysis

Aim = to find the most desirable combination of assets (portfolio) to hold.

Markowitz 1959: *Portfolio selection - Efficient diversification of investments*
then Sharpe.

1 Mean variance approach for investors' preferences

We consider a financial market on one period.
There are no transactions costs and no taxes. Assets are divisible: one can hold α assets, with $\alpha \in \mathbb{R}$ ($\alpha < 0$ means short selling, see below).

General assumption for the investors' behaviour:

- investors structure their portfolios in order to maximize the expected utility of their final wealth $E[U(W)]$, with $U(\cdot)$ utility function of the investor, defined on his final wealth W.

- their utility increases with wealth and they are risk averse:
U strictly increasing and concave ($U' > 0,\ U'' < 0$).

In the general case, a Taylor expansion around $\mathbb{E}\,(W)$ (or everywhere for an analytical function) implies $U(W) = U(\mathbb{E}\,(W))$

$$+U'(\mathbb{E}\,(W))(W - \mathbb{E}\,(W)) + \frac{1}{2}U''(\mathbb{E}\,(W))(W - \mathbb{E}\,(W))^2 + \sum_{n=3}^{+\infty} \frac{1}{n!}U^{(n)}(\mathbb{E}\,(W))(W - \mathbb{E}\,(W))^n.$$

Taking the expectation, we get:

$$\mathbb{E}\,[U(W)] = U(\mathbb{E}\,(W)) + \frac{1}{2}U"(\mathbb{E}\,(W))\mathbb{E}\,[(W - \mathbb{E}\,(W))^2] + \sum_{n=3}^{+\infty} \frac{1}{n!}U^{(n)}(\mathbb{E}\,(W))\mathbb{E}\,[(W - \mathbb{E}\,(W))^n].$$

Then for U strictly increasing and concave, the investor has a preference for expected value and aversion for variance. But the next moments are involved as well.

The mean variance approach relies on the assumption that investors select assets according to the expectation and variance of their final wealth or, more often, according to the expectation and variance of the (rate of) return on their wealth or investment ($W = W_0(1+R)$, function of $W \Leftrightarrow$ function of R). It is then restrictive.

Reminder about the first moments of a distribution

To assess an investment having a return R, we look at the following moments:

⋆ 1st moment: expectation $\mathbb{E}(R)$ = measure of the average return.
For R discrete with $\{r_i\}_i$ its possible values, $\mathbb{E}(R) = \sum_i p_i r_i$.

⋆ 2nd (central) moment: variance $V(R) = \mathbb{E}([R - \mathbb{E}(R)]^2)$ ($= \sum_i p_i [r_i - \mathbb{E}(R)]^2$ for R discrete).
It is a measure of the dispersion.

The standard deviation $\sigma(R) = \sqrt{V(R)}$ can also be used ("volatility"' of the asset). Both are risk measures.

Note that $V(R)$ is a satisfying measure of risk only when the probability distribution of the returns are symmetric: then it is fine to take into account gains in the risk measure.
Otherwise other measures can be used: semi-variance, Value-at-Risk...

⋆ 3rd (standardized) moment: skewness $\mathbb{E}[(\frac{R-\mathbb{E}(R)}{\sigma(R)})^3]$.
Skewness is a measure of symmetry, or more precisely, the lack of symmetry.
A distribution is symmetric if both the left and right sides are the same relative to the center point ($R - \mathbb{E}(R)$ and $-(R - \mathbb{E}(R))$ have the same law).

Skewness = 0 for a symmetric distribution (indeed the density of the centered variable, f, is then an even function, so for any $n \in \mathbb{N}$, $x \mapsto x^{2n+1} f(x)$ is an odd function and $\int x^{2n+1} f(x)dx = 0$), in particular for a Gaussian variable.

A skewness > 0 means a distribution skewed to the right,
i.e. the right tail is long relative to the left tail.
In finance, positively skewed returns will be preferred.

⋆ 4th (standardized) moment: kurtosis $\mathbb{E}[(\frac{R-\mathbb{E}(R)}{\sigma(R)})^4]$.
Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution.
It is worth 3 for R Gaussian. Then the excess kurtosis, Kurtosis−3, is computed.

Distributions with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Distributions with low kurtosis tend to have a flat top near the mean rather than a sharp peak. A uniform distribution would be the extreme case.
In finance, returns with lower kurtosis will be preferred.

Mean-variance preferences assumption means that investors' preferences can be described by a function that depends only on the expected return and return variance (as soon as these quantities are finite). In particular, investors do not care about skewness and kurtosis.

Precisely, it is assumed that for any investor, we have, with U his utility function and W_0 his initial wealth:

(A) | $\mathbb{E}(U(W)) = \mathbb{E}(U(W_0(1+R))) = f(\mathbb{E}(R), V(R))$ for a given function f (depending on the investor) satisfying: for $(x,y) \in \mathbb{R} \times \mathbb{R}^+$, $\frac{\partial f}{\partial x}(x,y) > 0$ and $\frac{\partial f}{\partial y}(x,y) < 0$.

This assumption is satisfied in the two following cases:

First justification: investors have Quadratic utility functions: $U(W) = aW - bW^2$, with $a, b > 0$. Note that such a function can be used only for values of wealth satisfying $W < \frac{a}{2b}$ a.s., for the utility function to be an increasing function of the wealth.

Exercise 1. Prove that a Quadratic Utility Function satisfies **(A)**.

Quadratic utility is considered implausible because:

- assets' returns have to be bounded.

- it assumes that investors are equally averse to deviations above the mean as they are to deviations below the mean (but this is implied by assumption (A), so not considered as a drawback in our framework).

- quadratic utility implies an absolute risk aversion (ARA) increasing with wealth.

$\Rightarrow$ second approach is generally preferred.

Second justification: the assets' returns are assumed to constitute a Gaussian vector (then any linear combination of these returns is Gaussian).
(Also, note that W Gaussian $\Leftrightarrow R$ Gaussian).

Indeed
1. a Gaussian distribution is fully characterized by mean / variance (any moment with order of at least 3 can be written in terms of the first 2 moments).
2. the assumption implies that the portfolios returns are Gaussian. Then $\mathbb{E}(U(W))$ is a function only of the first 2 moments.

3. Then, for any utility function U positively monotone and concave, the investor has preference for expected return and aversion for variance (or for standard deviation, which is equivalent).
See Exercise **2**.

Limits of this approach: huge limit: distributions are not Gaussian!

Conventional assumption: stock markets behave according to a random Gaussian or "normal" distribution. See for example Fama, Eugene F. (1965), "Random Walks In Stock Market Prices". Financial Analysts Journal.
But due to its rigid shape, the Gaussian distribution lacks the flexibility to capture crucial features that are presented in financial return data. It not only underestimates the probability of large movements but also that of very small movements; and it overestimates the frequency of intermediate movements.

But clearly, the concern is with the larger losses. Large movements in prices (i.e. crashes) are much more common than what would be predicted with a normal distribution.
Looking at the Dow Jones index roughly over the last 80 years, we observe on average a daily loss of more than 3% approximately every four months. Under Gaussian assumption, one expects such losses only about once every 13 months. A drop of 6% or more happened about every three years on average and not every 175,000 years as the Gaussian assumption implies. Losses of more than 9% occurred about once every 17 years, not once in a period that is about 25,000 times longer than the age of our universe.

(also: the Gaussian model would predict that a crash like the one that occurred in 1987 would occur only once every 1087 years. Empirical observations, however, give evidence to show that such crashes can possibly occur once every 38 years).

2 Mean variance portfolio theory

Under the previous assumption, determine the portfolios that will be preferred to others.

2.1 Risk of a portfolio

W_0 initial wealth (assumed positive).
R_i = return of asset i, for $1 \leq i \leq N$.
x_i = fraction of the portfolio (i.e. proportion of wealth) invested in asset i (no assumption on the sign of x_i).
We require the investor to be fully invested then $\sum_i x_i = 1$.
The amount $x_i W_0$ is invested in asset i, the gain/loss is: $x_i W_0 R_i$.
Total for the portfolio = $\sum_i x_i W_0 R_i$, corresponds to the (rate of) return $R_X = \sum_i x_i R_i$, ie:

Return on a portfolio of assets = weighted average of return on the individual assets.

Expected return on the portfolio: $\boxed{\mathbb{E}\left(R_X\right) = \sum_i x_i \mathbb{E}\left(R_i\right)}$

Exercise **4.1** Do we get an equivalent relationship if x_i is the number of asset i in the portfolio?

Variance of the portfolio return:

• **2 assets**
$V(R_X) = V(x_1 R_1 + x_2 R_2) = \mathbb{E}\left([x_1 R_1 + x_2 R_2 - \mathbb{E}\left(x_1 R_1 + x_2 R_2\right)]^2\right)$
$= E\left(\{x_1[R_1 - \mathbb{E}\left(R_1\right)] + x_2[R_2 - \mathbb{E}\left(R_2\right)]\}^2\right)$
$V(R_i) = \sigma_i^2$, with σ_i standard deviation of R_i,
$\mathbb{E}\left([R_1 - \mathbb{E}\left(R_1\right)][R_2 - \mathbb{E}\left(R_2\right)]\right) = Cov(R_1, R_2)$ denoted by σ_{12}.

Then $V(R_X) = x_1^2 \sigma_1^2 + 2 x_1 x_2 \sigma_{12} + x_2^2 \sigma_2^2$.

The covariance is a measure of how returns on assets move together.
$\sigma_{12} > 0$ when good (or bad) outcomes for each asset occur together.

• N **assets** (note that the results of this part do not require $\sum_i x_i = 1$)

$\boxed{V(R_X) = \sum_i x_i^2\sigma_i^2 + \sum_{i\neq j} x_i x_j \sigma_{ij}}$ with $\sigma_{ij} = Cov(R_i, R_j)$.

Also $V(R) = \sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j \sigma_{ij}$ where $\sigma_{ii} = \sigma_i^2$.

N variance terms, $N(N-1)$ covariance terms.

Easier with matrix writing.

Reminder (covariance properties): the covariance is a bilinear form:

$Cov(R, aR_1 + bR_2) = \mathbb{E}\,[R(aR_1 + bR_2)] - \mathbb{E}\,(R)\mathbb{E}\,(aR_1 + bR_2)$

$= a[\mathbb{E}\,(RR_1) - \mathbb{E}\,(R)\mathbb{E}\,(R_1)] + b[\mathbb{E}\,(RR_2) - \mathbb{E}\,(R)\mathbb{E}\,(R_2)] = a\ Cov(R, R_1) + b\ Cov(R, R_2)$.

It is symmetric: $Cov(R_1, R_2) = Cov(R_2, R_1)$, positive semi-definite:

$Cov(R, R) = V(R) \geq 0$, and

$Cov(R, R) = 0$ implies that R is a constant random variable (no randomness).

In particular, it satisfies the Cauchy-Schwartz inequality: $Cov(X, Y)^2 \leq V(X)V(Y)$,

which means that a correlation is always between -1 and 1.

Let $E_R = \begin{pmatrix} E(R_1) \\ ... \\ E(R_N) \end{pmatrix}$ and $\Sigma = \begin{pmatrix} \sigma_{11} & ... & \sigma_{1N} \\ ... & ... & ... \\ \sigma_{N1} & ... & \sigma_{NN} \end{pmatrix}$, returns covariance matrix for the N assets.

A portfolio is given by its weights $X = \begin{pmatrix} x_1 \\ ... \\ x_N \end{pmatrix}$. We have $R_X = {}^tX \begin{pmatrix} R_1 \\ ... \\ R_N \end{pmatrix}$ and $\boxed{\mathbb{E}\,(R_X) = {}^tXE_R}$.

For $1 \leq i \leq N$, $Cov(R_i, R_X) = \sum_j x_j Cov(R_i, R_j) = \sum_j \sigma_{ij} x_j$ then $\begin{pmatrix} Cov(R_1, R_X) \\ ... \\ Cov(R_N, R_X) \end{pmatrix} = \Sigma X$.

For another portfolio given by its weights $Y = \begin{pmatrix} y_1 \\ ... \\ y_N \end{pmatrix}$ (y_i = proportion of wealth invested in asset i), we have: $Cov(R_Y, R_X) = \sum_i y_i Cov(R_i, R_X)$ then $\boxed{Cov(R_Y, R_X) = {}^tY\Sigma X}$.

Note: 1. this result holds even if $\sum_i x_i$ is not equal to 1.

2. The proof could simply use the reminder at the end of the on of exercise 4, with $R = \begin{pmatrix} R_1 \\ ... \\ R_N \end{pmatrix}$:

$Cov(R_X, R_Y) = Cov({}^tXR, {}^tYR) = {}^tXCov(R, R)Y = {}^tY\Sigma X$.

Therefore $\boxed{V(R_X) = {}^tX\Sigma X}$.

The covariance matrix is symmetric and positive semi-definite, i.e. ${}^tX\Sigma X \geq 0$ for any X.

If Σ is invertible (its lines are linearly independent, Σ is then positive definite), $(X, Y) \mapsto {}^tX\Sigma Y$ is a symmetric positive definite bilinear form (scalar product): ${}^tX\Sigma X > 0$ for all non-zero vectors X.

Note: ${}^tX\Sigma X = 0 \Leftrightarrow V(\sum_{i=1}^N x_iR_i) = 0 \Leftrightarrow \sum_{i=1}^N x_iR_i$ non-random $\Leftrightarrow X$ risk-free portfolio.

Then Σ invertible means that the returns are "linearly independent", meaning that it is not possible to combine the assets in a risk-free portfolio. All possible portfolios built with these assets are risky.

2.2 Diversification

We assume $N \geq 2$ and no short sales (for any portfolio, $\forall i,\ x_i \geq 0$).

⋆ When all assets are independent: $V(R_X) = \sum_i x_i^2\sigma_i^2$.

For a portfolio containing at least 2 different assets, $\forall i,\ x_i < 1$ thus $x_i^2 < x_i$, $V(R_X) < \sum_i x_i\sigma_i^2$, which is the average variance of the returns.

= diversification; portfolio less risky than single assets.

If equal amounts are invested in each asset: $x_i = \frac{1}{N}$,

$V(R_X) = \sum_i \frac{1}{N^2}\sigma_i^2 = \frac{1}{N}\sum_i \frac{\sigma_i^2}{N} = \frac{1}{N}\times$ average variance of the returns.

For N large, $V(R)$ approaches 0.

⋆ Non-independent assets with $x_i = \frac{1}{N}$:

$$V(R_X) = \frac{1}{N}\underbrace{\sum_i \frac{\sigma_i^2}{N}}_{average\ variance} + \frac{N-1}{N}\underbrace{\sum_{i\neq j}\frac{\sigma_{ij}}{N(N-1)}}_{average\ covariance} = \frac{1}{N}\overline{\sigma_i^2} + \frac{N-1}{N}\overline{\sigma_{ij}}.$$

For N large, 1st term approaches 0, 2nd term approaches the average covariance.

The individual risk of securities can be diversified away, but the contribution to the total risk caused by the covariance terms cannot be diversified away.

Effect of diversification on portfolio risk: the minimum variance is obtained for very large portfolios and is equal to the average covariance between all stocks in the population. For an index where N is large, one can get close to the average covariance.

How is $V(R_X) = {}^tX\Sigma X$ modified by changing x_i? (without assuming $\sum_i x_i = 1$)

def: gradient of a function f defined on $\mathbb{R}^N$ (denoted by ∇f): vector of its N partial derivatives.

The gradient of $X \mapsto {}^tX\Sigma X$ is $2\Sigma X$.

Indeed: can be proven by full calculus (expanding the matrix product, see exercise 4) or:
the gradient of $X \mapsto {}^tAX$ if A is a column vector with N lines, is A.

The gradient of $(X,Y) \mapsto {}^tY\Sigma X$ is ${}^t\Sigma Y$ w.r.t. X and ΣX w.r.t. Y (since ${}^tY\Sigma X = {}^tX\,{}^t\Sigma Y$).
Then the gradient of $X \mapsto {}^tX\Sigma X$ w.r.t. X is ${}^t\Sigma X + \Sigma X = 2\Sigma X$.

We saw $\Sigma X = \begin{pmatrix} Cov(R_1, R_X) \\ ... \\ Cov(R_N, R_X) \end{pmatrix}$, then for $1 \leq i \leq N$, $\dfrac{\partial V(R_X)}{\partial x_i} = 2Cov(R_i, R_X)$.

Therefore the effect on $V(R_X)$ of increasing x_i is proportional to $Cov(R_i, R_X)$:
if this covariance is positive, $V(R_X)$ increases.

$Cov(R_i, R_X)$ is a measure of the marginal contribution of asset i to the portfolio risk.

2.3 Portfolios built with two risky assets

We represent in the mean/standard deviation plane (expected return plotted against standard deviation of the return, i.e. volatility) the investment opportunity set, i.e. the portfolios that can be built when combining 2 risky assets. This set depends on whether short selling is allowed or not.

Def (for any number of assets) : Global minimum variance portfolio (**GMV**)
= the portfolio that has the lowest risk of any feasible portfolio.

Same notations as before. We assume for example $0 < \sigma_1 \leq \sigma_2$ (both assets are risky).
We have $x_2 = 1 - x_1$. To simplify the writing below, we set $x = x_1$ ie $X = \begin{pmatrix} x \\ 1-x \end{pmatrix}$.

The portfolio's expected return is obtained as:
$\mathbb{E}(R_X) = x\mathbb{E}(R_1) + (1-x)\mathbb{E}(R_2) = x\mu_1 + (1-x)\mu_2$, where $\mu_i = \mathbb{E}(R_i)$.

With ρ denoting the correlation between R_1 and R_2: $\rho = \dfrac{\sigma_{12}}{\sigma_1\sigma_2} \in [-1, 1]$,
the portfolio return variance is: $V(R_X) = x^2\sigma_1^2 + 2x(1-x)\rho\sigma_1\sigma_2 + (1-x)^2\sigma_2^2$.

2.3.1 Short sales not allowed

For $x \in]0, 1[$, $V(R_X)$ is an increasing function of ρ.

• $\rho = 1$: perfectly correlated assets
$V(R_X) = [x\sigma_1 + (1-x)\sigma_2]^2$ or $\sigma(R_X) = x\sigma_1 + (1-x)\sigma_2$.
All the combinations of the 2 assets lie, in risk and return space,
on a straight line between asset 1 and asset 2. Indeed:

$$\begin{pmatrix} \mathbb{E}(R_X) \\ \sigma(R_X) \end{pmatrix} = x \begin{pmatrix} \mu_1 \\ \sigma_1 \end{pmatrix} + (1-x) \begin{pmatrix} \mu_2 \\ \sigma_2 \end{pmatrix}$$

(note: the equation of the line can be computed when $\sigma_1 < \sigma_2$: $x = \frac{\sigma(R_X) - \sigma_2}{\sigma_1 - \sigma_2}$
then $\mathbb{E}(R_X) = \frac{\sigma(R_X)-\sigma_2}{\sigma_1-\sigma_2}\mu_1 + (1 - \frac{\sigma(R_X)-\sigma_2}{\sigma_1-\sigma_2})\mu_2 = c^{ste} + c^{ste}\ \sigma(R_X)$).

There is no reduction in risk from purchasing both assets. Asset 1 is the GMV portfolio.

Note that in the particular case where $\sigma_1 = \sigma_2$ and $\rho = 1$, we get for any combination of the 2 assets: $V(R_X) = \sigma_1^2[x^2 + 2x(1-x) + (1-x)^2] = \sigma_1^2$. All the portfolios have the same risk, they are all a GMV portfolio, the preferred portfolio will be asset i such that $\mu_i = \max(\mu_1, \mu_2)$.

• $\rho = -1$: assets move together but in exactly opposite directions.
$\sigma(R_X) = |x\sigma_1 - (1-x)\sigma_2|$ risk smaller than for $\rho = 1$.

A portfolio with $x = \frac{\sigma_2}{\sigma_1+\sigma_2} \in]0, 1[$ will have zero risk (involves positive investment in both securities).

If $x \geq \frac{\sigma_2}{\sigma_1+\sigma_2}$, $\sigma(R_X) = x\sigma_1 - (1-x)\sigma_2$ and

$$\begin{pmatrix} \mathbb{E}(R_X) \\ \sigma(R_X) \end{pmatrix} = x \begin{pmatrix} \mu_1 \\ \sigma_1 \end{pmatrix} + (1-x) \begin{pmatrix} \mu_2 \\ -\sigma_2 \end{pmatrix}.$$

If $x \leq \frac{\sigma_2}{\sigma_1+\sigma_2}$, $\sigma(R_X) = -x\sigma_1 + (1-x)\sigma_2$ and

$$\begin{pmatrix} \mathbb{E}(R_X) \\ \sigma(R_X) \end{pmatrix} = x \begin{pmatrix} \mu_1 \\ -\sigma_1 \end{pmatrix} + (1-x) \begin{pmatrix} \mu_2 \\ \sigma_2 \end{pmatrix}.$$

Then $x = \frac{\sigma(R_X)+\sigma_2}{\sigma_1+\sigma_2}$ or $x = \frac{\sigma(R_X)-\sigma_2}{\sigma_1+\sigma_2}$ and $\mathbb{E}(R_X) = x\mu_1 + (1-x)\mu_2 = c^{ste} + c^{ste}\ \sigma(R_X)$ in both cases.

The opportunity set is constituted of 2 straight lines crossing at the point where $\sigma(R_X) = 0$ (i.e. $x = \frac{\sigma_2}{\sigma_1+\sigma_2}$), which corresponds to the GMV portfolio.

In that case, diversification is possible.

• Intermediate values of ρ: $\sigma(R_X)$ is increasing with ρ.
It reaches its lowest value for $\rho = -1$ and its highest value for $\rho = 1$.
In the plane, the 2 curves represent the limits which all portfolios of these two securities will lie within for intermediate values of ρ.
A given value of x corresponds to an horizontal line ($\mathbb{E}(R_X)$ given).

For a given ρ, there is a value of x that minimizes the portfolio risk:
$V(R_X) = f(x)$ where $f(x) = x^2\sigma_1^2 + 2x(1-x)\rho\sigma_1\sigma_2 + (1-x)^2\sigma_2^2 = \alpha x^2 - 2x\sigma_2(\sigma_2 - \rho\sigma_1) + \sigma_2^2$
with $\alpha = \sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2 \geq \sigma_1^2 - 2\sigma_1\sigma_2 + \sigma_2^2 = (\sigma_1 - \sigma_2)^2 \geq 0$,
$\alpha = 0$ iff $\rho = 1$ and $\sigma_1 = \sigma_2$, we exclude that case (treated above). Then $\alpha > 0$.

We can prove (see Exercise **5.**) that f has a unique minimum, at $x^* = \dfrac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}$.

We have $\rho\sigma_1 < \sigma_2$ (indeed $\rho\sigma_1 \leq \sigma_2$ and $\rho\sigma_1 = \sigma_2$ iff $\rho = 1$ and $\sigma_1 = \sigma_2$, excluded) then $x^* > 0$.

And $x^* \leq 1 \Leftrightarrow \sigma_1 - \rho\sigma_2 \geq 0$, then:

- For $\rho < \frac{\sigma_1}{\sigma_2}$, $x^* \in]0,1[$, i.e. the risk is minimum at a portfolio that is not one of the 2 assets (e.g. case $\rho = -1$ above). f has a strict minimum at x^*, then $f(x^*) < f(1) = \sigma_1^2$.

- For $\rho \geq \frac{\sigma_1}{\sigma_2}$, $f(x) \geq x^2\sigma_1^2 + 2x(1-x)\sigma_1^2 + (1-x)^2\sigma_2^2 \geq \sigma_1^2$, no portfolio less risky than asset 1 (e.g. case $\rho = 1$ above). Note that if $\rho = \frac{\sigma_1}{\sigma_2}$, $x^* = \frac{\sigma_2^2 - \sigma_1^2}{\sigma_1^2 - 2\sigma_1^2 + \sigma_2^2} = 1$.

Conclusion (relationship between expected return and standard deviation of the return for various correlation coefficients when short sales are not allowed):

- Combinations of 2 assets can never have more risk than the risk found on the straight line connecting the 2 assets in mean / standard deviation space (from $\sigma(R_X)$ increasing function of ρ). These combinations are on the left of that line.

- For $\rho < \frac{\sigma_1}{\sigma_2}$, the GMV portfolio is strictly less risky than the least risky asset; diversification is possible.

- For $\rho \geq \frac{\sigma_1}{\sigma_2}$, the risk on the portfolio can no longer be made less than the risk of the least risky asset in the portfolio.
The GMV portfolio is asset 1 (the least risky asset), no diversification effect.

- The lower (closer to -1) the correlation coefficient between assets (all other attributes held constant), the higher the payoff from diversification.

2.3.2 Short sales allowed

An investor can sell a security that he does not own (and buy it back later). Short selling involves taking a negative position in the security (equivalent to: borrow, sell, buy back to reimburse).
We assume that there are no transaction costs involved in this process.

Note that a portfolio where I sell n assets 2 and put the outcome in asset 1 is not considered (as W_0 would be 0, excluded).

Now we can have $x < 0$ (curve beyond asset 2)
or $x > 1$ (curve beyond asset 1).

⋆ No limit anymore to the expected return of feasible portfolios.
Indeed one can sell asset 1 (if it has the lowest expected return)
to invest more in asset 2 ($x < 0$).

⋆ The point x^* where f is minimum is now achievable by a portfolio, even if $\rho > \frac{\sigma_1}{\sigma_2}$ (then $x^* > 1$).
Diversification is then always possible. The GMV always corresponds to the minimum of f.

(see details in Exercise 6).

2.4 Delineating efficient portfolios for N risky assets

As for 2 assets, we represent the **opportunity set** (possible portfolios) in the mean/standard deviation plane and the subset of portfolios that will be preferred by all investors (**efficient portfolios**) under mean/variance assumption.
These sets depend if short selling and lending/borrowing (of funds) are allowed or not.

2.4.1 The shape of the opportunity set for N risky assets, no risk-free asset

= set of all possible combinations of assets in the mean/standard deviation space.
In the next section, we will analytically determine this set. Here, we say a few things about its shape.

The points on the frontier of this set will correspond to portfolios that have the smallest risk for a given expected return.

Th: The portion of the frontier that lies above (below) the GMV portfolio is concave (convex).
Proof: combinations of the GMV portfolio (point the most on the left) and another portfolio A on the frontier:

⋆ a portfolio A that has a higher return (and higher variance). Possible shapes?

a: NO. The combinations of GMV and A would be missing:
all combinations of GMV and A must lie either on the straight line
connecting GMV and A or above this straight line.

b: NO. Here all portfolios have less risk than the straight line
connecting the 2 assets, which is fine.
But, U and V are portfolios (combinations of GMV portfolio and asset A)
then all combinations of U and V must lie either on the straight line
connecting U and V or above this straight line.

c = only possible shape: concave.

⋆ a portfolio A with a lower return (and higher variance).
Same arguments (the combinations of GMV and A
are on the left of the line connecting these 2 points):

Efficient portfolios

An investor will prefer a lower risk for the same return, and a bigger return for the same risk.

A portfolio is said to be **efficient** if

there is no portfolio with the same risk and a higher expected return

there is no portfolio with the same expected return and a lower risk.

Other portfolios are called "dominated".

a. The efficient frontier with no short sales

B preferred to A (higher return with same risk)

C preferred to A

Therefore the set of efficient portfolios cannot
include interior portfolios (dominated portfolios).

D preferred to E

This is true for every other portfolio as we move up the outer shell from B to C.

Point C cannot be eliminated since there is no portfolio that has less risk for the same return or more return for the same risk. C is the GMV portfolio (it has the lowest risk of any feasible portfolio).
F is on the outer shell, but D has less risk for the same return.
As we move up the outer shell curve from F, all portfolios are dominated until we come to B.
B represents that portfolio (usually a single security) that offers the highest expected return of all portfolios.

Thus the efficient set (called the **efficient frontier**) consists of the envelope curve of all portfolios that lie between the GMV portfolio and the maximum return portfolio, i.e. the higher part of the curve (there can be linear segments between two perfectly correlated efficient portfolios).
The efficient frontier is concave.

b. The efficient frontier with short sales allowed

There is no limit anymore to the expected return of the portfolios one can build. Indeed one can sell securities with low expected returns and use the proceeds to buy securities with high expected returns.

The efficient set still starts with the GMV portfolio, but when
short sales are allowed it has no finite upper bound
(the curve goes beyond the maximum return portfolio).

2.4.2 The shape of efficient frontier when there is a risk-free asset

Until now, we have been dealing with portfolios of risky assets. We introduce now a risk-free asset into our portfolio possibility set. Risk-free lending and borrowing are possible.

Lending at a risk-free rate $\Leftrightarrow$ investing in an asset with a certain outcome.
Borrowing $\Leftrightarrow$ selling such a security short. This takes place at the risk-free rate (unique from an assumption of no arbitrage opportunity).

r certain (rate of) return on the risk-free asset. The standard deviation of this return is 0.
We assume that investors can lend and borrow unlimited amounts of funds at the risk-free rate.

1. The investor is interested in placing part of the funds in some portfolio A and either lending or borrowing.
Determine the geometric pattern of all combinations of (long) portfolio A and lending or borrowing.
$x \geq 0$ fraction of wealth invested in portfolio A. The investor can borrow at the risk-free rate and invest more than his initial funds in portfolio A. Then x can be greater than 1.

Expected return of the investment: $\mathbb{E}(R) = x\mathbb{E}(R_A) + (1-x)r = r + x[\mathbb{E}(R_A) - r]$.
Risk: $\sigma(R) = x\sigma(R_A)$.

Then $\mathbb{E}(R) = r + \frac{\mathbb{E}(R_A)-r}{\sigma(R_A)}\sigma(R)$,

i.e. all combinations of risk-free lending or borrowing with portfolio A
lie on a straight line in mean/standard deviation plan.

Intercept of the line on the return axis: r.
The line passes through the point $(\sigma_A, E(R_A))$. Left (right) of point A
= combinations of portfolio A and lending (borrowing).

2. changing A: e.g. B, combinations have greater return for the same risk

⇒ rotate the line as far as possible, until it becomes
tangent to the efficient frontier (for the set of risky assets).
there are no portfolios lying above the line passing through r and T.

All investors should hold the portfolio of risky assets T
(same for all investors, whatever their risk-aversion).

Some of these investors who are quite risk-averse would select a portfolio along the segment rT (combination of risk-free asset and T).
Others would hold portfolios on the half-line beyond T, borrowing funds and placing their original capital plus the borrowed funds in risky portfolio T.
Or T alone.

> All of these investors will hold risky portfolios with the exact composition of portfolio T. Then identifying T solves the portfolio problem. The ability to determine the optimum portfolio of risky assets without having to know anything about the investor is called the separation theorem.

Draw the efficient frontier under more restrictive assumptions about the ability of investors to borrow at the risk-free rate (exercise 7):

a. if they cannot borrow at all,
b. if they can lend at one rate, but must pay a higher rate r' to borrow.

2.5 Calculation of the efficient frontier when there is no risk-free asset

Assumptions:

- N risky assets, R_i = (rate of) return of asset i, $1 \le i \le N$. σ_i = standard deviation of R_i.

- The $\mathbb{E}(R_i)$ are not all equal (else, the best portfolio is just the asset i with the smallest σ_i).

- No risk-free asset. Then cash lending or borrowing is not permitted.

  In particular: · $\sigma_i > 0$ (really risky assets),

  · the covariance matrix of the N assets' returns, Σ, is invertible: it is not possible to combine the assets in a risk-free portfolio (possible portfolios are all risky).

  Σ is then symmetric positive definite.

- Short sales are allowed, i.e. x_i = fraction of the portfolio invested in asset i can be negative: if W_0 is the total wealth, $|x_i|W_0$ is the cash obtained from the sale of asset i, this cash will be invested in some other assets. We keep $\sum_{i=1}^N x_i = 1$ but x_i can have any sign: $x_i \in \mathbb{R}$.

2.5.1 Minimum variance frontier

= portfolios which minimize the return variance for a given expected return μ.

A portfolio given by its weights on the N assets, $X = (x_i)_{i=1...N}$, corresponds to a minimum variance portfolio iff it solves the optimization problem:

$$Min V(R_X) \text{ under the constraints } \mathbb{E}(R_X) = \mu \text{ and } \sum_{i=1}^N x_i = 1.$$

Reminder: for a portfolio X, we have $R_X = \sum_{i=1}^N x_i R_i = {}^tRX$,

then $\mathbb{E}(R_X) = {}^tE_R X$ where $E_R = \begin{pmatrix} \mathbb{E}(R_1) \\ ... \\ \mathbb{E}(R_N) \end{pmatrix}$ and $V(R_X) = {}^tX\Sigma X$ with $\Sigma = \begin{pmatrix} \sigma_{11} & ... & \sigma_{1N} \\ ... & ... & ... \\ \sigma_{N1} & ... & \sigma_{NN} \end{pmatrix}$.

Then we have to solve:

$$Min\, {}^tX\Sigma X \text{ under the constraints } {}^tE_R X = \mu \text{ and } {}^t\mathbf{1}X = 1, \text{ with } \mathbf{1} = \begin{pmatrix} 1 \\ ... \\ 1 \end{pmatrix}.$$

quadratic programming problem, solved by method of Lagrangian multipliers (standard method to find the extrema of a function subject to constraints).

The function $X \mapsto {}^tX\Sigma X$ is convex on $\mathbb{R}^N$, indeed the square matrix of its second-order partial derivatives (the covariance matrix 2Σ) is positive definite. Note also that our second assumption implies E_R not proportional to $\mathbf{1}$.

The Lagrangian is $\mathcal{L} = {}^tX\Sigma X + \lambda(\mu - {}^tE_R X) + \gamma(1 - {}^t\mathbf{1}X)$.

Necessary condition to have a minimum: the derivatives w.r.t. X, λ, and γ have to be null.

Note: $\frac{\partial \mathcal{L}}{\partial \lambda} = \frac{\partial \mathcal{L}}{\partial \gamma} = 0 \Longleftrightarrow$ the contraints are satisfied.
The problem (N equations, $p = 2$ constraints, N unknowns) becomes an unconstrained problem ($N + p$ equations, $N + p$ unknowns) when the multipliers are added in the variables.

The gradient of $X \mapsto {}^tX\Sigma X$ is $2\Sigma X$.

Necessary and sufficient conditions to have a global minimum:

$$2\Sigma X - \lambda E_R - \gamma \mathbf{1} = \mathbf{0} \quad , \quad \frac{\partial \mathcal{L}}{\partial \lambda} = \mu - {}^tE_R X = 0 \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial \gamma} = 1 - {}^t\mathbf{1}X = 0.$$

$$\left(\text{i.e. } \frac{\partial \mathcal{L}}{\partial x_i} = 2\sum_{j=1}^{N} x_j\sigma_{ij} - \lambda \mathbb{E}(R_i) - \gamma = 0 \text{ for } i = 1, ..., N, \qquad \mu - \sum_{i=1}^{N} x_i \mathbb{E}(R_i) = 0 \text{ and } 1 - \sum_{i=1}^{N} x_i = 0\right).$$

We get (with Σ^{-1} the inverse matrix of Σ): $2X = \lambda\Sigma^{-1}E_R + \gamma\Sigma^{-1}\mathbf{1}$,

thus $\mu = \frac{1}{2}(\lambda \underbrace{{}^tE_R\Sigma^{-1}E_R}_{b} + \gamma \underbrace{{}^tE_R\Sigma^{-1}\mathbf{1}}_{a})$ and $\frac{1}{2}(\lambda \underbrace{{}^t\mathbf{1}\Sigma^{-1}E_R}_{a} + \gamma \underbrace{{}^t\mathbf{1}\Sigma^{-1}\mathbf{1}}_{c}) = 1$,

with $a = {}^t\mathbf{1}\Sigma^{-1}E_R = {}^tE_R\Sigma^{-1}\mathbf{1}$, $b = {}^tE_R\Sigma^{-1}E_R$, $c = {}^t\mathbf{1}\Sigma^{-1}\mathbf{1}$.

Note that $\underline{b, c > 0}$ since Σ^{-1} is a symmetric positive definite matrix (and E_R cannot be the vector 0 as not proportional to $\mathbf{1}$).

Let $d = bc - a^2$. We have $\underline{d > 0}$, indeed $(X, Y) \mapsto {}^tX\Sigma^{-1}Y$ is a symmetric positive definite bilinear form (scalar product), then the Cauchy-Schwartz inequality applies:

$a^2 = ({}^tE_R\Sigma^{-1}\mathbf{1})^2 \leq {}^tE_R\Sigma^{-1}E_R \cdot {}^t\mathbf{1}\Sigma^{-1}\mathbf{1} = bc$ and not equal since E_R not proportional to $\mathbf{1}$.

We have: $2X = \lambda\Sigma^{-1}E_R + \gamma\Sigma^{-1}\mathbf{1}$, $\quad \mu = \frac{1}{2}(\lambda b + \gamma a)$ and $\frac{1}{2}(\lambda a + \gamma c) = 1$.

Therefore $\lambda = 2\frac{c\mu - a}{d}$ and $\gamma = 2\frac{b - a\mu}{d}$.

Then the composition of the frontier portfolio with an expected return equal to μ is given by:

$$\boxed{X = \frac{c\mu - a}{d}\Sigma^{-1}E_R + \frac{b - a\mu}{d}\Sigma^{-1}\mathbf{1}}$$

Any portfolio of the frontier can be represented by a vector X satisfying this equation, for $\mu = E(R_X)$. And any portfolio given by such a vector is a frontier portfolio:

> <u>Th</u>: The minimum variance frontier is composed of all portfolios $P(\mu), \mu \in \mathbb{R}$ such that, for a given expected return μ, the relative weights are:
>
> $X(\mu) = \frac{\mu}{d}\Sigma^{-1}(cE_R - a\mathbf{1}) + \frac{1}{d}\Sigma^{-1}(-aE_R + b\mathbf{1})$,
>
> with $a = {}^t\mathbf{1}\Sigma^{-1}E_R = {}^tE_R\Sigma^{-1}\mathbf{1}$, $b = {}^tE_R\Sigma^{-1}E_R$, $c = {}^t\mathbf{1}\Sigma^{-1}\mathbf{1}$ and $d = bc - a^2$.

[i.e. $a = \sum_{i=1}^{N}\sum_{j=1}^{N} v_{ij}\mathbb{E}(R_j)$, $\quad b = \sum_{i=1}^{N}\sum_{j=1}^{N} v_{ij}\mathbb{E}(R_i)\mathbb{E}(R_j)$, $\quad c = \sum_{i=1}^{N}\sum_{j=1}^{N} v_{ij}$

$x_i = \frac{\mu}{bc - a^2}\sum_{j=1}^{N} v_{ij}(c\mathbb{E}(R_j) - a) + \frac{1}{bc - a^2}\sum_{j=1}^{N} v_{ij}(b - a\mathbb{E}(R_j))$ if $\Sigma^{-1} = (v_{ij})_{1 \leq i,j \leq N}$].

> <u>Th</u>: The minimum variance frontier is a parabola (hyperbola) when expected return is plotted against return variance (standard deviation).

Proof: variance of the return of a portfolio of the frontier, when $\mathbb{E}(R_X) = \mu$:

$$V(R_X) = {}^tX\Sigma X = \frac{1}{2}{}^tX(\lambda E_R + \gamma\mathbf{1}) = \frac{1}{2}(\lambda {}^tXE_R + \gamma) = \frac{1}{2}(\lambda\mu + \gamma) = \frac{c\mu - a}{d}\mu + \frac{b - a\mu}{d}.$$

Therefore $\boxed{V(R_X) = \frac{1}{d}[cE(R_X)^2 - 2aE(R_X) + b]}$, equation of a parabola.

Generally represented in the mean/standard deviation space, thus hyperbola (see tutorial, exercise 8).

Particular minimum variance portfolios $X(\mu) = \frac{c\mu - a}{d}\Sigma^{-1}E_R + \frac{b - a\mu}{d}\Sigma^{-1}\mathbf{1}$:

· The GMV portfolio has been defined as having the lowest risk of any feasible portfolio:

$\frac{\partial V(R_{X(\mu)})}{\partial \mu} = \frac{2}{d}(c\mu - a) = 0$ iff $\mu = \frac{a}{c}$, then $V(R_{GMV}) = \frac{1}{d}[c(\frac{a}{c})^2 - 2a\frac{a}{c} + b] = \frac{1}{c}$.

The GMV portfolio weights are: $X_{GMV} = X(\frac{a}{c}) = \frac{b - a\frac{a}{c}}{d}\Sigma^{-1}\mathbf{1} = \frac{1}{c}\Sigma^{-1}\mathbf{1}$.

· for $a \neq 0$: $X(\frac{b}{a}) = \frac{c\frac{b}{a} - a}{d}\Sigma^{-1}E_R = \frac{1}{a}\Sigma^{-1}E_R$.

Notes: 1. if $a > 0$, $\frac{b}{a} > \frac{a}{c}$, then $X(\frac{b}{a})$ is above the GMV portfolio. Else, it is below.

2. we have: $\forall \mu \in \mathbb{R}$, $X(\mu) = \frac{c\mu - a}{d}aX(\frac{b}{a}) + \frac{b - a\mu}{d}cX_{GMV}$: any minimum variance portfolio is the linear combination of the 2 minimum variance portfolios $X(\frac{b}{a})$ and X_{GMV}.

The parabola equation can be rewritten by taking the GMV portfolio as the origin:

$V(R_{X(\mu)}) - V(R_{GMV}) = \frac{1}{d}[c\mu^2 - 2a\mu + b] -$ same function at $\frac{a}{c}$.

Then $V(R_{X(\mu)}) - \frac{1}{c} = \frac{1}{d}\{c[\mu^2 - (\frac{a}{c})^2] - 2a[\mu - \frac{a}{c}]\} = \frac{1}{d}[\mu - \frac{a}{c}][c\mu + a - 2a] = \frac{c}{d}[\mu - \frac{a}{c}]^2$.

Therefore $\boxed{V(R_X) - \frac{1}{c} = \frac{c}{d}[E(R_X) - \frac{a}{c}]^2}$.

Covariance of the GMV portfolio return with the return of any combination of assets:

$Cov(R_{GMV}, R_X) = {}^tX_{GMV}\Sigma X = \frac{1}{c}\,{}^t(\Sigma^{-1}\mathbf{1})\Sigma X = \frac{1}{c}\,{}^t\mathbf{1}X = \frac{1}{c} = V(R_{GMV})$.

2.5.2 Efficient frontier

Th: The efficient frontier is the upper half of a parabola (hyperbola) when expected return is plotted against return variance (standard deviation). It is composed of all portfolios $P(\mu), \mu \geq \frac{a}{c}$ such that, for a given expected return μ, the relative weights are given by:

$X = \dfrac{\mu}{d}\Sigma^{-1}(cE_R - a\mathbf{1}) + \dfrac{1}{d}\Sigma^{-1}(-aE_R + b\mathbf{1})$, with a, b, c, d as before.

Equation of the efficient frontier: $V(R_X) = \frac{1}{d}[cE(R_X)^2 - 2aE(R_X) + b]$ for $\mathbb{E}(R_X) \geq \frac{a}{c}$.

upper half of a parabola starting from the point $(\frac{1}{c}, \frac{a}{c})$ (GMV), with a positive slope (or of a hyperbola starting from the point $(\frac{1}{\sqrt{c}}, \frac{a}{c})$ in the mean / standard deviation plane).

Ex: the GMV portfolio is efficient, and for $a > 0$, $X(\frac{b}{a})$ is efficient. If $a < 0$, $X(\frac{b}{a})$ is dominated.

Two-fund separation theorem (Black, 1972):
The efficient frontier can be generated from any 2 efficient portfolios.

i.e. if X_1, X_2 are 2 different efficient portfolios, any other efficient portfolio can be written $\alpha X_1 + (1 - \alpha)X_2$.

Proof: the same result holds in fact for the whole minimum variance frontier: it is generated from any 2 minimum variance (MV) portfolios. We start with that case.

A MV portfolio satisfies:

$X = \dfrac{E(R_X)}{d}\Sigma^{-1}(cE_R - a\mathbf{1}) + \dfrac{1}{d}\Sigma^{-1}(-aE_R + b\mathbf{1}) = E(R_X)G + H$ (linear in $E(R_X)$)

with $G = \frac{1}{d}\Sigma^{-1}(cE_R - a\mathbf{1})$ and $H = \frac{1}{d}\Sigma^{-1}(b\mathbf{1} - aE_R) \in \mathbb{R}^N$.

- Let X_1, X_2 be 2 different MV portfolios, then they have different expected returns (indeed one-to-one mapping $\mu \mapsto X(\mu)$ efficient). Then, any 3rd MV portfolio X_3 can be written as a linear combination of X_1 and X_2. Indeed:

there exists α such that $E(R_{X_3}) = \alpha E(R_{X_1}) + (1-\alpha)E(R_{X_2})$,

then $X_3 = E(R_{X_3})G + H = \alpha[E(R_{X_1})G + H] + (1-\alpha)[E(R_{X_2})G + H] = \alpha X_1 + (1-\alpha)X_2$.

Therefore the minimum variance frontier is: $\{\alpha X_1 + (1-\alpha)X_2 | \ \alpha_X \in \mathbb{R}\}$.

- Let X_1, X_2 be 2 different efficient portfolios, the MV frontier is still $\{\alpha X_1 + (1-\alpha)X_2 | \ \alpha_X \in \mathbb{R}\}$, but if we restrict the possible values for α, we get the efficient frontier:
if for example $E(R_{X_1}) < E(R_{X_2})$,
$\alpha X_1 + (1-\alpha)X_2$ is efficient iff $\alpha E(R_{X_1}) + (1-\alpha)E(R_{X_2}) \geq \frac{a}{c}$ ie $\alpha[E(R_{X_2}) - E(R_{X_1})] \leq E(R_{X_2}) - \frac{a}{c}$

The efficient frontier is then the set $\{\alpha X_1 + (1-\alpha)X_2 \mid \alpha \leq \frac{E(R_{X_2}) - \frac{a}{c}}{E(R_{X_2}) - E(R_{X_1})}\}$.

Consequence: you can combine 2 (or more) investment managers who have efficient portfolios, your portfolio will remain efficient (as soon as the proportions of both portfolios satisfy above inequality).

Remark: we saw that a portfolio X is efficient iff: $\Sigma X = \frac{cE(R_X)-a}{d} E_R + \frac{b-aE(R_X)}{d} \mathbf{1}$ and $E(R_X) \geq \frac{a}{c}$ (then the 1st coefficient is ≥ 0).

The next theorem states that such a linear relationship characterises the efficient portfolios.

> Th (Roll): Necessary and sufficient condition for a portfolio X to be efficient:
>
> $\exists \lambda_X, \gamma_X$ with $\lambda_X \geq 0$ s.t. $\Sigma X = \lambda_X E_R + \gamma_X \mathbf{1}$ (positive linear relationship between ΣX and E_R),
>
> i.e. $$\begin{pmatrix} Cov(R_1, R_X) \\ ... \\ Cov(R_N, R_X) \end{pmatrix} = \lambda_X \begin{pmatrix} \mathbb{E}(R_1) \\ ... \\ \mathbb{E}(R_N) \end{pmatrix} + \gamma_X \begin{pmatrix} 1 \\ ... \\ 1 \end{pmatrix} \qquad (N \text{ equations}).$$

Proof: portfolio given by its weights X. The vector of covariance $(Cov(R_i, R_X))_i$ is equal to ΣX. To be proved: ΣX is a linear combination of vectors E_R and $\mathbf{1}$.

1. If X is efficient, this is satisfied with $\lambda_X = \frac{cE(R_X)-a}{d} \geq 0$ (in fact > 0 except for the GMV portfolio).

2. if X satisfies $\Sigma X = \lambda_X E_R + \gamma_X \mathbf{1}$:

we get 2 conditions over λ_X and γ_X by computing ${}^t\mathbf{1}X$ and ${}^tE_R X$:

multiplying left by ${}^t\mathbf{1}\Sigma^{-1}$, we get ${}^t\mathbf{1}X = \lambda_X \, {}^t\mathbf{1}\Sigma^{-1}E_R + \gamma_X \, {}^t\mathbf{1}\Sigma^{-1}\mathbf{1}$, ie $1 = \lambda_X a + \gamma_X c$.

Then, first proof: multiplying left by ${}^tE_R\Sigma^{-1}$, we get

${}^tE_R X = \lambda_X \, {}^tE_R\Sigma^{-1}E_R + \gamma_X \, {}^tE_R\Sigma^{-1}\mathbf{1}$, ie $E(R_X) = \lambda_X b + \gamma_X a$

then $cE(R_X) = \lambda_X bc + (1 - \lambda_X a)a$ and $\lambda_X = \frac{cE(R_X)-a}{d}$, which implies $E(R_X) \geq \frac{a}{c}$,

while $\gamma_X = \frac{1}{c}[1 - a\frac{cE(R_X)-a}{d}] = \frac{1}{cd}[bc - a^2 - acE(R_X) + a^2] = \frac{1}{d}[b - aE(R_X)]$.

We get $\Sigma X = \frac{cE(R_X)-a}{d} E_R + \frac{b-aE(R_X)}{d} \mathbf{1}$ with $E(R_X) \geq \frac{a}{c}$. Then X is efficient.

or, second proof (but only if $a \neq 0$): $X = \lambda_X \Sigma^{-1}E_R + \gamma_X \Sigma^{-1}\mathbf{1} = \lambda_X a X(\frac{b}{a}) + \gamma_X c X_{GMV}$ is a linear combination of 2 MV portfolios, then it is a MV portfolio.

Moreover, $E(R_X) = \lambda_X a\frac{b}{a} + \gamma_X c\frac{a}{c} = \lambda_X a\frac{b}{a} + (1-\lambda_X a)\frac{a}{c} = \frac{a}{c} + \lambda_X(b - \frac{a^2}{c}) = \frac{a}{c} + \lambda_X\frac{d}{c} \geq \frac{a}{c}$.
Therefore X is efficient.

We saw that $Cov(R_i, R_X)$ is a measure of the marginal contribution of asset i to the portfolio risk (end of 2.2). Here we see that a portfolio is mean-variance efficient iff the marginal contribution of each asset to the portfolio risk is a linear function of the assets' expected returns.

We now rewrite the linear relationship of Roll theorem. The coefficients will be expressed in terms of an orthogonal portfolio.

2.5.3 Orthogonal portfolio

Th: To any minimum variance portfolio X different from GMV can be associated another minimum variance portfolio $X^\perp$ that is not correlated to X ($Cov(R_X, R_{X^\perp}) = 0$), called *orthogonal portfolio* of X (it is orthogonal for the metrics $(Y,Z) \mapsto {}^tY\Sigma Z$).
$X^\perp$ is on the opposite part of the frontier and satisfies $E(R_{X^\perp}) = \dfrac{aE(R_X) - b}{cE(R_X) - a}$.
We have also $(X^\perp)^\perp = X$.

Proof: any minimum variance portfolio $X(\mu)$ satisfies $X(\mu) = \frac{1}{d}\Sigma^{-1}([c\mu - a]E_R + [b - a\mu]\mathbf{1})$.
We set $\mu_X = E(R_X)$.
For any other portfolio Y (combination of the N risky assets, not necessarily a MV portfolio), let $\mu_Y = E(R_Y)$. We have:

$Cov(R_X, R_Y) = {}^tX\Sigma Y = \frac{1}{d}([c\mu_X - a]\,{}^tE_R + [b - a\mu_X]\,{}^t\mathbf{1})Y$

$= \frac{1}{d}([c\mu_X - a]\mu_Y + [b - a\mu_X]) = \frac{c}{d}([\mu_X - \frac{a}{c}]\mu_Y + \frac{b}{c} - \frac{a}{c}\mu_X) = \frac{c}{d}(\mu_X - \frac{a}{c})(\mu_Y - \frac{a}{c}) + \frac{1}{c}$.

Then $Cov(R_X, R_Y) = 0 \Leftrightarrow (\mu_X - \frac{a}{c})(\mu_Y - \frac{a}{c}) = -\frac{d}{c^2}$

By assumption $\mu_X \neq \frac{a}{c}$ then $Cov(R_X, R_Y) = 0 \;\Leftrightarrow\; (c\mu_X - a)(c\mu_Y - a) = a^2 - bc \;\Leftrightarrow\; \mu_Y = \frac{a\mu_X - b}{c\mu_X - a}$.

The portfolios on the half-line corresponding to this value of the expected return are not correlated to X.

We define $X^\perp$ as the minimum variance portfolio Y among these portfolios.

We have $Cov(R_X, R_{X^\perp}) = 0$ and $\mathbb{E}\,(R_{X^\perp}) = \frac{a\mu_X - b}{c\mu_X - a}$.

Note that if $\mu_X > \frac{a}{c}$, then $\mu_{X^\perp} < \frac{a}{c}$ (from $\frac{c}{d}(\mu_X - \frac{a}{c})(\mu_{X^\perp} - \frac{a}{c}) < 0$) and inversely if X is a frontier portfolio that is dominated.

Note: we get for X on the upper part of the frontier ($\mu_X \geq \frac{a}{c}$) and any other portfolio Y s.t. $\mu_Y \geq \frac{a}{c}$:

$Cov(R_X, R_Y) = \frac{c}{d}(\mu_X - \frac{a}{c})(\mu_Y - \frac{a}{c}) + \frac{1}{c} > 0$:
X is positively correlated with any portfolio with an expected return larger that $\frac{a}{c}$,
in particular all the efficient portfolios are positively correlated.

Geometric construction of the orthogonal portfolio:

1. The line $(GMV\ X)$ in the variance/mean plane cuts the mean-axis at point $(0, \mu_{X^\perp})$.

Proof: see tutorial (exercise 9).

2. In the $(\sigma(\cdot), E(\cdot))$ plane, the tangent to the frontier at X cuts the mean-axis at point $(0, \mu_{X^\perp})$.

Proof: the equation of the upper part of the frontier (efficient frontier) is $\sqrt{\sigma^2 - \frac{1}{c}} = \sqrt{\frac{c}{d}}(\mu - \frac{a}{c})$ i.e. $\mu = \frac{a}{c} + g(\sigma)$ for $\sigma \geq \frac{1}{\sqrt{c}}$, where $g(\sigma) = \sqrt{\frac{d}{c}(\sigma^2 - \frac{1}{c})}$.

Tangent to the frontier at $(\sigma, \frac{a}{c} + g(\sigma))$: line $(\sigma, \frac{a}{c} + g(\sigma)) + \lambda(1, g'(\sigma)),\ \lambda \in \mathbb{R}$,

cuts ordinate axis for $\lambda = -\sigma$, at a point $(0, \mu')$ s.t. $\mu' = \frac{a}{c} + g(\sigma) - \sigma g'(\sigma)$

then $(\mu - \frac{a}{c})(\mu' - \frac{a}{c}) = g(\sigma)(\mu' - \frac{a}{c}) = g(\sigma)^2 - \sigma g(\sigma)g'(\sigma) = \frac{d}{c}(\sigma^2 - \frac{1}{c}) - \frac{d}{c}\sigma^2$.

We get $(\mu - \frac{a}{c})(\mu' - \frac{a}{c}) = -\frac{d}{c^2}$ which proves that $\mu' = \mu_{X^\perp}$.

Note that the two previous constructions hold because X is not the GMV portfolio.

Th (Roll bis): Necessary and sufficient condition for a portfolio X to belong to the efficient frontier (except for GMV): $\exists$ a portfolio Y non-correlated with X s.t.

$$E_R = E(R_Y)\mathbf{1} + \frac{E(R_X) - E(R_Y)}{V(R_X)} \Sigma X \qquad \text{with } \mathbb{E}(R_X) - \mathbb{E}(R_Y) > 0.$$

i.e. $\forall 1 \leq i \leq N, \quad E(R_i) = E(R_Y) + \dfrac{E(R_X) - E(R_Y)}{V(R_X)} Cov(R_i, R_X)$.

Proof: 1. If Y exists, we have a positive linear relationship as in Roll theorem, then X is efficient.

2. For X efficient, $\Sigma X = \frac{c\mu_X - a}{d} E_R + \frac{b - a\mu_X}{d}\mathbf{1}$,

thus $E_R = \frac{a\mu_X - b}{c\mu_X - a}\mathbf{1} + \frac{d}{c\mu_X - a}\Sigma X = \mu_{X^\perp}\mathbf{1} + \frac{d}{c\mu_X - a}\Sigma X$.

But $V(R_X) = \frac{1}{d}[c\mu_X^2 - 2a\mu_X + b] = \frac{1}{d}[(c\mu_X - a)\mu_X + b - a\mu_X] = \frac{1}{d}[(c\mu_X - a)(\mu_X - \mu_{X^\perp})]$

Then $\frac{d}{c\mu_X - a} = \frac{\mu_X - \mu_{X^\perp}}{V(R_X)}$, hence the result, with $Y = X^\perp$.

$\forall i,\ E(R_i) = \mu_{X^\perp} + \frac{\mu_X - \mu_{X^\perp}}{\sigma_X^2} Cov(R_i, R_X)$. And $\mu_X - \mu_{X^\perp} > 0$ from X efficient $\neq$ GMV.

= positive linear relationship between the expected return of asset i and its marginal contribution to the risk of an efficient portfolio other than GMV.

2.6 Calculation of the efficient frontier when there is a risk-free asset

Same assumptions:

- N risky assets $(\forall i,\ \sigma_i > 0)$,

the $\mathbb{E}(R_i)$ are not all equal, and the covariance matrix of the N assets' returns, Σ, is invertible.

- Short sales are allowed (no constraints on weights sign).

Plus:

- There exists a risk-free asset: asset 0. Cash lending or borrowing is permitted, at a same risk-free rate r.

A portfolio is given by its weights $(x_i)_{i=0,\dots,N}$ where $x_i \in \mathbb{R}$ is the proportion of wealth invested in asset i. In fact, the vector of weights on the risky assets is fully determining the portfolio as $x_0 = 1 - \sum_{i=1}^{N} x_i = 1 - {}^tX\mathbf{1}$.

Then, for $X = \begin{pmatrix} x_1 \\ \dots \\ x_N \end{pmatrix}$, we can denote by R_X the return of the portfolio $\begin{pmatrix} x_0 \\ X \end{pmatrix}$.

It should be denoted by $R_{\begin{pmatrix} x_0 \\ X \end{pmatrix}}$ but no risk of confusion as X determines the portfolio.

We have $R_X = (1 - {}^tX\mathbf{1})r + {}^tX \begin{pmatrix} R_1 \\ \dots \\ R_N \end{pmatrix}$, then $\mathbb{E}(R_X) = (1 - {}^tX\mathbf{1})r + {}^tXE_R$

(note that $\mathbb{E}(R_X) - r = {}^tX(E_R - r\mathbf{1})$) and $V(R_X) = \sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j \sigma_{ij} = {}^tX\Sigma X$.

2.6.1 Minimum variance frontier

$X = (x_i)_{i=1\dots N}$ corresponds to a minimum variance portfolio iff it minimizes the return variance for a given expected return μ, i.e. it solves the optimization problem:

$Min\, {}^tX\Sigma X$ under the constraint $(1 - {}^tX\mathbf{1})r + {}^tXE_R = \mu$.

Lagrangian $\mathcal{L} = {}^tX\Sigma X + \lambda\left[\mu - r - {}^tX(E_R - r\mathbf{1})\right]$

Since the covariance matrix Σ is invertible, below conditions are necessary and sufficient to get a global minimum: $\nabla\mathcal{L} = 2\Sigma X - \lambda(E_R - r\mathbf{1}) = 0$
and $\frac{\partial\mathcal{L}}{\partial\lambda} = \mu - r - {}^tX(E_R - r\mathbf{1}) = 0$. Therefore we have:

Th: The minimum variance frontier is composed of all portfolios $P(\mu), \mu \in \mathbb{R}$ with the following weights for a given expected return μ (a, b, c as before):

$$X(\mu) = \frac{\mu - r}{f^2}\Sigma^{-1}(E_R - r\mathbf{1}), \text{ with } f = \sqrt{cr^2 - 2ar + b} \quad \text{and} \quad x_0 = 1 - \sum_{i=1}^{N} x_i.$$

Proof: we get $X = \frac{\lambda}{2}\Sigma^{-1}(E_R - r\mathbf{1})$ and $\mu - r = {}^tX(E_R - r\mathbf{1}) = \frac{\lambda}{2}\,{}^t(E_R - r\mathbf{1})\Sigma^{-1}(E_R - r\mathbf{1})$.

As Σ^{-1} is positive definite and E_R and $\mathbf{1}$ not proportional, ${}^t(E_R - r\mathbf{1})\Sigma^{-1}(E_R - r\mathbf{1})$ is positive, we denote it by f^2, with $f > 0$.
We have $f^2 = \underbrace{{}^tE_R\Sigma^{-1}E_R}_{b} - r\underbrace{{}^t\mathbf{1}\Sigma^{-1}E_R}_{a} - r\underbrace{{}^tE_R\Sigma^{-1}\mathbf{1}}_{a} + r^2\underbrace{{}^t\mathbf{1}\Sigma^{-1}\mathbf{1}}_{c} = b - 2ar + cr^2$.

Then $\mu - r = \frac{\lambda}{2}f^2$. It gives: $\frac{\lambda}{2} = \frac{\mu - r}{f^2}$, then X as in the theorem.

Note that the $X(\mu)$ for $\mu \in \mathbb{R}$ are all proportional to vector $\Sigma^{-1}(E_R - r\mathbf{1})$.

Th: Again, the minimum variance frontier is a parabola when expected return is plotted against return variance.
Proof: for X as in the theorem, we have:

$$V(R_X) = {}^tX\Sigma X = \frac{\mu - r}{f^2}\,{}^tX(E_R - r\mathbf{1}) = \frac{\mu - r}{f^2}(E(R_X) - r) = \frac{(\mu_X - r)^2}{f^2}.$$

For $\mu_X = r$, we get the portfolio minimizing the variance: the risk-free asset.

Any other portfolio is risky.
The parabola is tangent at the ordinate axis at the corresponding point $(0, r)$.

Then, for any frontier portfolio, we have $\sigma_X = \begin{cases} \dfrac{\mu_X - r}{f} & \text{if } \mu_X \geq r \\ \dfrac{-\mu_X + r}{f} & \text{else} \end{cases}$

Therefore, in the space standard deviation / expected return,
the minimum variance frontier is made of 2 half-lines with slopes $\pm f$.

Obviously, only the upper half-line (with the positive slope) corresponds to efficient portfolios. The portfolios on the other line are dominated.

2.6.2 Tangent portfolio and separation theorem

Is there any minimum variance portfolio containing risky assets only?
Reminder: the portfolios constituted of risky assets are delimited by a hyperbola in the mean / standard deviation plane (the least risky portfolio satisfies $(E(R_X), \sigma(R_X) = (\frac{a}{c}, \frac{1}{\sqrt{c}}))$.
This hyperbola is on the right of the 2 half-lines (which are the minimum variance frontier when including a risk-free asset). Indeed, the existence of the risk-free asset leads to a lower risk at any expected return level (and to a higher expected return at any risk level).

A minimum variance portfolio containing only risky assets would satisfy:
${}^tX\mathbf{1} = 1$ (i.e. $x_0 = 0$: all the wealth invested in risky assets) and $X = \dfrac{\mu - r}{f^2}\Sigma^{-1}(E_R - r\mathbf{1})$ (frontier portolio).
Then ${}^tX\mathbf{1} = \dfrac{\mu - r}{f^2}\,{}^t(E_R - r\mathbf{1})\Sigma^{-1}\mathbf{1} = \dfrac{\mu - r}{f^2}({}^tE_R\Sigma^{-1}\mathbf{1} - r\,{}^t\mathbf{1}\Sigma^{-1}\mathbf{1}) = \dfrac{\mu - r}{f^2}(a - cr)$ has to be equal to 1.

We assume $r \neq \frac{a}{c}$. Then ${}^tX\mathbf{1} = 1$ iff $\mathbb{E}\,(R_X) - r = \dfrac{f^2}{a - cr}$ i.e. $\mathbb{E}\,(R_X) = r + \dfrac{cr^2 - 2ar + b}{a - cr} = \dfrac{ar - b}{cr - a}$.

So there exists such a portfolio, $\mathbf{T}$, with $\mathbb{E}\,(R_T) = \dfrac{ar - b}{cr - a}$ and $\sigma(R_T) = \dfrac{|\mu_T - r|}{f} = \dfrac{f}{|a - cr|}$, and it corresponds to the following allocation X_T on the risky assets:

$$X_T = \frac{E(R_T) - r}{f^2}\Sigma^{-1}(E_R - r\mathbf{1}) = \frac{1}{a - cr}\Sigma^{-1}(E_R - r\mathbf{1}) \text{ since } \mathbb{E}\,(R_T) - r = \frac{f^2}{a - cr}.$$

Note that $\mathbb{E}\,(R_T) > r$ iff $a - cr > 0$.

Also, T cannot be the GMV portfolio, indeed, we would have $\mathbb{E}\,(R_T) = \dfrac{a}{c}$ then $c(ar - b) = a(cr - a)$ and $d = 0$ (which is excluded by the assumption that the $\mathbb{E}\,(R_i)$ are not all equal).

T minimizes the variance among the portfolios built on the $N + 1$ assets. It belongs to the set of portfolios built on the N risky assets only, so it minimizes the variance on this set as well. That means that it belongs to the hyperbola.
T is the intersection of the hyperbola with one of the two half-lines mentioned above: the upper line iff T is efficient, which happens iff $\mathbb{E}\,(R_T) > r$, i.e. iff $r < \frac{a}{c}$. Else, if $r > \frac{a}{c}$, T is on the lower line.

Since $\mathbb{E}\,(R_T) = \dfrac{ar - b}{cr - a}$, portfolio $T^\perp$ (exists as $T \neq$ GMV) has an expected return equal to r.

That means the line $(0, r) - T$ in the $(E(\cdot), \sigma(\cdot))$ plane is tangent to the hyperbola at T (see 2nd Geometric construction of the orthogonal portfolio), as announced in 2.4.2.

Two-fund separation theorem (Tobin, 1958)
As before, the minimum variance frontier can be generated from any 2 minimum variance portfolios. But we have a more precise result here:

Th: When $r \neq \frac{a}{c}$, any minimum variance portfolio can be written as a linear combination of the risk-free asset and the tangent portfolio T.

Proof: any minimum variance portfolio satisfies: $X = \frac{\mu_X - r}{f^2}\Sigma^{-1}(E_R - r\mathbf{1})$.
But $X_T = \frac{1}{a - cr}\Sigma^{-1}(E_R - r\mathbf{1})$, then $X = \frac{\mu_X - r}{f^2}(a - cr)X_T = \alpha_X X_T$, with $\alpha_X \in \mathbb{R}$ and ${}^tX\mathbf{1} = 1 - x_0 = \alpha_X\, {}^tX_T\mathbf{1} = \alpha_X$.

Thus the portfolio is the combination of α_X portfolio T and $1 - \alpha_X$ risk-free asset.
Precisely, we have: $\begin{pmatrix} x_0 \\ X \end{pmatrix} = \alpha_X \begin{pmatrix} 0 \\ X_T \end{pmatrix} + (1 - \alpha_X)\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, with $\alpha_X = \frac{V(R_X)}{E(R_X) - r}(a - cr)$.

Note that: $V(R_X) = {}^tX\Sigma X = \alpha_X^2\, {}^tX_T\Sigma X_T = \alpha_X^2 V(R_T)$.

2.6.3 Efficient frontier

Th: The efficient frontier is composed of all portfolios $P(\mu), \mu \geq r$ with the following weights for a given expected return μ (with a, b, c as before):

$$X(\mu) = \frac{\mu - r}{f^2}\Sigma^{-1}(E_R - r\mathbf{1}), \text{ with } f = \sqrt{cr^2 - 2ar + b} \quad \text{and} \quad x_0 = 1 - {}^tX\mathbf{1}.$$

The efficient frontier starts at point $(0, r)$ (risk-free asset).
For the other portfolios, the higher the variance the higher the expected return.

Note: we get, for any efficient portfolio: $\frac{E(R_X) - r}{\sigma(R_X)} = f$, this is known as the Sharpe ratio (see exercise 16). We can see here that this ratio is the same for all efficient portfolios.

When $r \neq \frac{a}{c}$, any efficient portfolio is the combination of $1 - \alpha_X$ risk-free asset and α_X portfolio T. The combination of risky assets will then be the same for all investors. But the quantities of risk-free asset / portfolio T depend on each investor's risk aversion.

• If $r < \frac{a}{c}$, T satisfies $\mu_T = \mathbb{E}(R_T) > r$, T is on the upper line and $\sigma_T = \frac{\mu_T - r}{f} = \frac{f}{a - cr}$.
The portfolio T is efficient and the efficient frontier is tangent to the hyperbola at T.

Efficient frontier = $\{\alpha_X$ portf T + $(1 - \alpha_X)$ risk-free asset $| \ 0 \leq \alpha_X\}$.
Indeed X efficient iff $\mathbb{E}(R_X) - r \geq 0$ but $\mathbb{E}(R_X) - r = \alpha_X(\mathbb{E}(R_T) - r)$ has the sign of α_X.
$\alpha_X \geq 1$ means borrowing at r to buy more of T.
The other line (dominated portfolios) corresponds to short sales of T,
with product of sale invested at rate r.

• If $r > \frac{a}{c}$, T satisfies $\mu_T < r$ (below $(0, r)$), then $\sigma_T = \frac{-\mu_T + r}{f} = \frac{f}{cr - a}$.
The portfolio T is dominated.
The line with a negative slope is tangent to the hyperbola.
Intersection at T s.t. $(\mu_T, \sigma_T) = (\frac{b - ar}{a - cr}, \frac{f}{cr - a})$.

Efficient frontier = $\{\alpha_X$ portf T + $(1-\alpha_X)$ risk-free asset $\mid \alpha_X \le 0\}$.
All efficient portfolios are obtained by short selling T, to invest at rate r
(combination with positive weight α_X on T are dominated).

• If $r = \frac{a}{c}$, no intersection between the 2 straight lines and the hyperbola.
Then $f^2 = cr^2 - 2ar + b = \frac{a^2}{c} - 2\frac{a^2}{c} + b = \frac{bc-a^2}{c} = \frac{d}{c}$. The 2 straight lines are given by:

$$\sigma_X = \begin{cases} \sqrt{\frac{c}{d}}[\mu_X - r] & \text{if } \mu_X \ge r \\ \sqrt{\frac{c}{d}}[-\mu_X + r] & \text{else} \end{cases}$$

i.e. $\mu_X = \frac{a}{c} \pm \sqrt{\frac{d}{c}}\sigma_X$. Exercise **15.1**: Prove that the 2 lines constituting the MV frontier correspond to the asymptotes of the hyperbola.

Any minimum variance portfolio satisfies:

$$X = c\frac{\mu_X - r}{d}\Sigma^{-1}(E_R - \frac{a}{c}\mathbf{1}) = (\mu_X - r)X_A, \text{ with } X_A = \frac{c}{d}\Sigma^{-1}(E_R - \frac{a}{c}\mathbf{1}).$$

We denote by A the combination of risky assets corresponding to the vector of weights X_A.

Exercise **15.2**: Prove that investing in A costs nothing.

15.3: Prove that the portfolio $\begin{pmatrix} 1 \\ X_A \end{pmatrix}$ is efficient and compute its return and its variance.

15.4: Prove that the efficient frontier is $\{\alpha_X$ portf $\begin{pmatrix} 1 \\ X_A \end{pmatrix}$ + $(1-\alpha_X)$ risk-free asset $\mid \alpha_X \ge 0\}$.

Any minimum variance portfolio is obtained by investing all wealth in the risk-free asset and holding more or less of A.

Note that we will see that at market equilibrium, necessarily $r < \frac{a}{c}$.

> <u>Th</u> (Roll): Necessary and sufficient condition for a risky portfolio X to belong to the efficient frontier: $\forall i, \quad E(R_i) = r + \frac{E(R_X) - r}{V(R_X)}Cov(R_i, R_X) \quad$ with $\mathbb{E}(R_X) - r > 0$.

Note: above result holds, whatever the relative positions of r and $\frac{a}{c}$.

Proof: in the previous version, $X^\perp$ was needed. Here, any efficient portfolio can be expressed in terms of T or A. Only $T^\perp$ is used.

1. X efficient $\Rightarrow \Sigma X = \frac{\mu_X - r}{f^2}(E_R - r\mathbf{1}) \Rightarrow \forall i,\ E(R_i) = r + \frac{f^2}{\mu_X - r}Cov(R_i, R_X)$ while $V(R_X) = [\frac{\mu_X - r}{f}]^2$.

2. if relation holds, $\Sigma X = \frac{V(R_X)}{E(R_X)-r}(E_R - r\mathbf{1})$. But $\Sigma^{-1}(E_R - r\mathbf{1})$ is equal to $(a-cr)X_T$ or $\frac{d}{c}X_A$:
   If $r \ne \frac{a}{c}$, $X = \frac{V(R_X)}{E(R_X)-r}(a-cr)X_T$, remaining wealth is in the risk-free asset $\Rightarrow X$ efficient (combination of T and risk-free asset).
   If $r = \frac{a}{c}$, $X = \frac{V(R_X)}{E(R_X)-r}\frac{d}{c}X_A$, remaining in risk-free asset $\Rightarrow$ efficient. Indeed ${}^tX1 = 0$, then, the portfolio corresponding to X is: $\begin{pmatrix} 1 \\ X \end{pmatrix} = \alpha_X \begin{pmatrix} 1 \\ X_A \end{pmatrix} + (1-\alpha_X)\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, with $\alpha_X = \frac{V(R_X)}{E(R_X)-r}\frac{d}{c}$.

3 Limits in the practical implementation

The mean-variance portfolio analysis has several limits when implemented.

Amount of required input data

The model requires a huge number of estimates to fill the covariance matrix.
Example: $N = 100 \Rightarrow 100$ estimates of expected returns, 100 estimates of variance, $\frac{N(N-1)}{2} = 4950$ estimates of covariances.
Also, estimation of the covariance coefficients can lead to inconsistent values (the matrix has to be definitive positive).
In the next chapter, the estimation of the covariance matrix is simplified.

Also portfolio managers typically have reliable returns forecasts for only a small subset of assets.
Quality forecasts for the variance-covariance structure is even harder to obtain.
The ex post performance of the resulting weighting schemes depends heavily on the quality of the input data, especially the vector of expected returns. Historical returns are bad proxies for future expected returns.

Extreme portfolio weights

Optimized allocations tend to include large short positions, which cannot be justified by the portfolio manager in a client's portfolio. They also tend to overweight assets with large expected returns, negative correlations and small variances, but intuitively, assets with extreme returns tend to be the most affected by estimation error.

Instability of portfolios weights

Optimal allocations are particularly sensitive to changes in expected returns.
Small changes in input variables can cause dramatic changes in the weighting schemes.
Remedies: - data cleaning (delete outliers, i.e. very extreme observations with very small probability,
- resampling techniques, shrinkage estimators techniques,
- introduce constraints to avoid extreme weights (may come from client-specific or legal restrictions).
Give better results.

Chapter II Capital Asset Pricing Model

The preceding chapter has been concerned with how an individual, acting upon a set of estimates, could select an optimal portfolio. If investors act as prescribed, we should be able to determine how the aggregate of investors will behave, and how prices and returns at which markets will clear are set: construction of general equilibrium models.
The CAPM describes relationships between the prices of assets at market equilibrium.

1 Equilibrium definition and first consequences

1.1 Assumptions

We consider the market on one period (single-period investment horizon) with the following assumptions:
for investors:
· the investors have mean variance preferences
· homogeneous expectations: the investors make all the same forecasts for mean and variance of returns.

on the financial market:

· N risky assets with same assumptions as in the previous chapter: no transactions costs, no taxes, assets are divisible, the expected returns of the assets are not all equal and the covariance matrix of the N assets' returns, Σ, is invertible.

· Short selling is allowed.

· The market is at equilibrium, defined as:
- each investor holds an optimal portfolio (individual equilibrium)
- demand = supply: all existing assets are owned by someone, i.e. no excess demand or supply (then observed price is an equilibrium price).

Then each investor holds an efficient portfolio and Chapter 1 results apply: the efficient frontier is known (half-line if there exists a risk-free asset, else upper half of an hyperbola).

1.2 Market portfolio

Def: the Market portfolio is a portfolio containing every risky assets available to investors in amounts proportional to their market values (i.e. $x_i = \frac{\text{capitalisation of asset } i}{\text{total market capitalisation}}$, capitalisation=asset quantity $\times$ asset price). We denote it by M.

Th: At market equilibrium, the Market portfolio is efficient.

Proof: Because all existing assets are owned by someone, the Market portfolio is a combination of all investors portfolios.
All investors see the same efficient frontier and all choose an efficient portfolio.
Then the Market portfolio is a combination of efficient portfolios.
Thus it is efficient (from the separation theorem, OK in both cases: with or without a risk-free asset).

I.e. market equilibrium (demand = supply) allows us to identify a particular portfolio on the efficient frontier.

1.3 Efficient frontier at equilibrium when there exists a risk-free asset

We assume that there exists a risk-free asset (a government bond is considered a risk-free investment). Then the market participants can borrow and lend money at the same risk-free rate of interest r, with no limit.

In the $(\sigma(\cdot), E(\cdot))$ plane, the efficient frontier is a half-line: investors all choose a portfolio that is a combination of the risk-free asset and of a common risky portfolio T (or A if $r = \frac{a}{c}$).

Th: $\boxed{\text{At equilibrium, } r < \dfrac{a}{c}\,.\ \text{ The market portfolio coincides with the tangent portfolio: } M = T}$

Proof: If $r = \frac{a}{c}$, all investors hold the combination of risky assets A and some risk-free asset.

Thus global demand of (i.e. wealth invested in) risky assets = 0, cannot give a market equilibrium as soon as net supply of these assets is positive.

Then $r \neq \frac{a}{c}$. All investors hold a combination of the risk-free asset and of portfolio T. Then portfolio T will be exactly the portfolio constituted of all risky assets, i.e. M (all existing assets are owned by someone).
This implies $r < \frac{a}{c}$ as M, then T, is efficient. Also, if we had $r > \frac{a}{c}$, all investors, shorting T, would hold negative quantities (in value) of M. Inconsistent with market equilibrium. Then assets prices will adjust s.t. $r < \frac{a}{c} = \frac{{}^tE_R\Sigma^{-1}\mathbf{1}}{{}^t\mathbf{1}\Sigma^{-1}\mathbf{1}}$ for the market to be at equilibrium.

So the efficient portfolios are the combinations of the risk-free asset and M.
Note that $E(R_M) > r$.

Conclusion: If investors all see the same efficient frontier and all choose an efficient portfolio, their portfolios are all on the half straight line starting from the risk-free asset and crossing the Market portfolio, with positive slope, called the Capital Market Line. The higher the risk, the higher the expected return. Dominated portfolios are below this line (e.g. individual assets).

This CML is constituted of all portfolios X satisfying, with x (≥ 0) invested in portfolio M, $1-x$ in risk-free asset:

$$(1-x)\begin{pmatrix}0\\ r\end{pmatrix} + x\begin{pmatrix}\sigma(R_M)\\ E(R_M)\end{pmatrix} = \begin{pmatrix}\sigma(R_X)\\ E(R_X)\end{pmatrix}. \quad \text{Thus } \boxed{\mathbb{E}\,(R_X) = r + \frac{E(R_M)-r}{\sigma(R_M)}\sigma(R_X)}$$

equation of the efficient frontier, common to all investors

(in chapter 1, we found $\mathbb{E}\,(R_X) = r + f\sigma(R_X)$ and $\sigma_T = \frac{\mu_T - r}{f}$, ie as $T = M$, $f = \frac{E(R_M)-r}{\sigma(R_M)}$, consistent).

Then the expected return of an efficient portfolio is the sum of r and $\mathbb{E}\,(R_X) - r$ which is an additional income for bearing the risk (risk premium):

risk premium by risk unit $= \dfrac{E(R_M)-r}{\sigma(R_M)}$, denoted by Π_M.
(depends on the average risk aversion of all market participants)

= price of risk = slope of the CML

= additional return required by investors
to bear an additional unit of risk (measured by the variance)
when investing in an efficient portfolio.

2 CAPM

The Capital Asset Pricing Model (CAPM), which was developed by Sharpe (1964), Lintner (1965), and Mossin (1966), is one of the cornerstones of modern finance. The CAPM is founded on Markowitz's mean-variance framework and on the assumptions of homogeneous expectations and no limitations on short-selling. Based on these (and other) assumptions, the CAPM derives a simple linear relation between risk and return. Sharpe, Markowitz and Miller jointly received in 1990 the Nobel Prize in Economics for this contribution to the field of financial economics.
Black (1972) formulated the "Zero-beta Capital Asset Pricing Model", which is the CAPM when there exists no risk-free asset (our first case below).

2.1 Zero-beta Capital Asset Pricing Model (no risk-free asset)

The market portfolio does not coincide with GMV (M=GMV is possible only if all investors choose that portfolio, it can be shown that that implies that all assets have the same expected return, and that is excluded here).

Th: At market equilibrium, there is a linear relationship between expected returns of the assets. Indeed, the expected return of any asset i is determined by:

$$\mathbb{E}(R_i) = E(R_{M^\perp}) + \beta_i[E(R_M) - E(R_{M^\perp})] \text{ with } \beta_i = \frac{Cov(R_i, R_M)}{V(R_M)} \text{ and } \mathbb{E}(R_M) - E(R_{M^\perp}) > 0.$$

$R_{M^\perp}$ is the return of a portfolio having a 0 covariance with the market portfolio.

Proof: direct consequence of Roll theorem: M is efficient (at market equilibrium) $\neq$ GMV, then there exists a portfolio $M^\perp$ non-correlated with M such that:

$$\forall i, \quad E(R_i) = E(R_{M^\perp}) + \frac{E(R_M)-E(R_{M^\perp})}{V(R_M)} Cov(R_i, R_M) \quad \text{and } \mathbb{E}(R_M) - E(R_{M^\perp}) > 0.$$

$\beta_i = \dfrac{Cov(R_i, R_M)}{V(R_M)}$ is called the beta of the asset.

β_i is proportional to $Cov(R_i, R_M)$, which measures the marginal contribution of asset i to the total risk of the market portfolio (cf end of 2.2 p5).

We have $\beta_M = 1$ and $\beta_{M^\perp} = 0$ (hence $M^\perp$ is a zero-beta portfolio).

The CAPM says that at market equilibrium, the expected return of an asset =
expected return of a zero-beta portfolio + a premium
where premium = return premium of the market ($\mathbb{E}(R_M) - E(R_{M^\perp})$)
$\times$ term proportional to the asset contribution to the total risk of the market portfolio.

The higher the beta, the higher the expected return at equilibrium.
The expected return of an asset depends on its covariance with the market portfolio.

We draw the expected return as a function of beta.

At market equilibrium, all traded assets are on a line.
The line is called the Security Market Line.
It goes through the market portfolio (whose beta is 1),
intersection with ordinate axis =
expected return of the zero-beta portfolio.

An important assumption behind the Zero-Beta CAPM is that short-sales are possible. To obtain zero-beta portfolios, we typically would have to short sell some assets. If there are short-sale constraints the Zero-Beta CAPM fails to hold.

2.2 CAPM when there is a risk-free asset

= more classical form of CAPM (Sharpe 1964, Lintner 1965, and Mossin 1968).

2.2.1 Formulation of CAPM, as a consequence of our previous results

Th: When there is a risk-free asset, at market equilibrium, the expected return of asset i is given
by: $\mathbb{E}(R_i) = r + \beta_i[E(R_M) - r]$ with $\beta_i = \dfrac{Cov(R_i, R_M)}{V(R_M)}$, beta of the asset.

Proof: direct consequence of the last Roll theorem as M is efficient and is not the GMV portfolio (as $T \neq GMV$).

The result can be written $\mathbb{E}(R_i) - r = \dfrac{E(R_M) - r}{\sigma(R_M)}\beta_i\sigma(R_M) = \pi_M\beta_i\sigma(R_M)$
= risk premium of asset i = additional return required by investors to bear the risk.
At market equilibrium, the risk premium on an individual asset is proportional to its covariance with the market portfolio (or to its beta, or to its contribution to the total risk of the market portfolio).

Expected return as a function of beta: Security Market Line.
Goes through the market portfolio,
intersection with y-axis = r.

At market equilibrium, all traded assets are on this SML line.

The higher the beta, the higher the expected return at equilibrium. beta=1 for the market.
Depending of β_i value, the market risk is amplified or softened.

Same equation than the CML, but holds for any asset, not only efficient portfolios:
CML: for X efficient, $\mathbb{E}(R_X) = r + \pi_M\sigma(R_X)$ where $\pi_M = \frac{E(R_M)-r}{\sigma(R_M)}$, risk premium by unit of risk.

SML: for any traded asset i, $\mathbb{E}(R_i) = r + \pi_M\beta_i\sigma(R_M)$.

Then if asset i is efficient, i.e. on both lines, we get $\sigma(R_i) = \beta_i\sigma(R_M)$. We come back to this later.

2.2.2 2nd interpretation of CAPM: linear regression of the asset i return on the market portfolio return

Looking at real data, a link can be observed between individual asset movements and whole market movements: eg, when the whole market goes up, each particular asset tends to go up.

This influence can be studied on the past moves series for market/ asset, through a linear regression of R_i on R_M.

Reminder:
→ line through a cloud of N points $(x_t, y_t)_t$, fitted using the least squares method:

Find (a, b) solving $Min \sum_{t=1}^{N} [y_t - (a + bx_t)]^2$.

$\frac{\partial}{\partial a} = 0$ iff $(\mathbf{1}) \sum_t [y_t - (a + bx_t)] = 0$ $\quad$ $\frac{\partial}{\partial b} = 0$ iff $(\mathbf{2}) \sum_t x_t [y_t - (a + bx_t)] = 0$

$$\begin{cases} aN + b\sum_t x_t = \sum_t y_t \\ a\sum_t x_t + b\sum_t x_t^2 = \sum_t x_t y_t \end{cases} \quad \text{then } b = \frac{\begin{vmatrix} N & \sum_t y_t \\ \sum_t x_t & \sum_t x_t y_t \end{vmatrix}}{\begin{vmatrix} N & \sum_t x_t \\ \sum_t x_t & \sum_t x_t^2 \end{vmatrix}} = \frac{N\sum_t x_t y_t - \sum_t x_t \sum_t y_t}{N\sum_t x_t^2 - (\sum_t x_t)^2}$$

Note that the residual terms (ε_t) defined by $y_t = a + bx_t + \varepsilon_t$
satisfy from equation **(1)**: $\sum_t \varepsilon_t = 0$ and from equation **(2)**: $\sum_t x_t \varepsilon_t = 0$.
This justifies the assumptions chosen in the linear regression below.

The smaller $\sum_t \varepsilon_t^2$ the better the line explains the set of data points.

→ linear regression of a variable Y on another variable X:

Assumption: $Y = a + bX + \varepsilon$, where ε is an error term s.t. $\mathbb{E}(\varepsilon) = Cov(\varepsilon, X) = 0$.

$Cov(Y - (a + bX), X) = 0$ implies $Cov(Y, X) = b\ Cov(X, X)$

then $b = \frac{Cov(X,Y)}{V(X)}$, while a is obtained as $a = \mathbb{E}(Y) - b\mathbb{E}(X)$ (called the y-intercept or intercept).

Note that as $Cov(\varepsilon, X) = 0$, we have $V(Y) = V(a + bX) + V(\varepsilon)$.
The regression is satisfying when $V(\varepsilon)$ is small compared to $V(a + bX)$, i.e. when $\frac{V(a+bX)}{V(Y)}$ is close to 1. The quality of the regression is then measured by the coefficient of determination:

$$R^2 = \frac{V(a + bX)}{V(Y)} = \frac{b^2 V(X)}{V(Y)} = \frac{Cov(X,Y)^2 V(X)}{V(X)^2 V(Y)} = \underline{Corr(X,Y)^2}.$$

R^2 measures the quality of the approximation by the line, i.e. how precisely Y values are explained by X values.

On observed data:
b is estimated by the empirical values of $\frac{Cov(X,Y)}{V(X)}$ (consistent with our first computation).

We apply this to $X = R_M$ and $Y = R_i$, assuming $R_i = a + bR_M + \varepsilon$, with additional properties for ε. The coefficient b is then obtained as:

$$b = \frac{Cov(R_M, R_i)}{V(R_M)}$$

which is what we obtained for β_i. The beta you get from Sharpe's derivation of equilibrium prices (CAPM) is essentially the same beta you obtain from doing a least-squares regression of the return of asset i on the return of the market portfolio.

But in the first approach we have assumptions of equilibrium, R_M and R_i are supposed to be future returns on the next period..., while the second approach is the result of a simple regression and past data are used.

In both cases, the beta will be estimated on past data, assuming that the distribution of (R_M, R_i) is stable over time (including their link of course).

The regression line of R_i on R_M is called the characteristic line of asset i. α_i is the intercept. Equation: $R_i = \alpha_i + \beta_i R_M$ (represents the relationship between the market return and the asset return, computed on past data, each series assumed to be observations of a same variable).

Example of characteristic line: fund NFMAX compared to Russell 3000 (good proxy of the U.S. market portfolio as large index representing 98% of the U.S. capitalisation), the weekly returns between May 2005 and July 2009 are used.
The resulting line is $R_i = 0.05\% + 126\%\ R_M$ i.e. $\beta = 1.26$.

Comparing the two approaches:

$$R_i = \alpha_i + \beta_i R_M + \varepsilon_i \Rightarrow \mathbb{E}(R_i) = \alpha_i + \beta_i E(R_M), \text{ to compare to } \mathbb{E}(R_i) = r + \beta_i[E(R_M) - r].$$

α_i should be equal to $r(1 - \beta_i)$ if the returns distributions are stationary and the series used for the regression is long enough.

On real data: if for example $\alpha_i > r(1 - \beta_i)$, that means that during regression period, stock did better than expected.
The difference between α_i and $r(1 - \beta_i)$ is called Jensen's alpha and provides a measure of whether the investment in question earned a return greater than or less than its required return, given both market performance and risk (see also Chapter III).

For instance, a firm that earned 15% during a period, when firms with similar betas earned 12%, will have earned an excess return of 3%; its intercept will also exceed $r(1 - \beta_i)$ by 3%.

A statistic usually computed is the standard error of the beta estimate. The slope of the regression, like any statistical estimate, may be different from the true value; and the standard error reveals just how much error there could be in the estimate. The standard error can also be used to arrive at confidence intervals for the true beta value from the slope estimate.

2.2.3 Usual model: "Market model"

also known as Single-Index or Diagonal model (Sharpe, 1963)
= empirical version of the CAPM.

Market model assumptions:
- for each asset: $R_i = \alpha_i + \beta_i R_M + \varepsilon_i$, with $\mathbb{E}(\varepsilon_i) = 0$, $Cov(R_M, \varepsilon_i) = 0$.
- the (ε_i) are all independent (uncorrelated is sufficient).

The investors have mean variance preferences (they will choose their portfolios according to the expected return and the return variance).

That is, asset moves depend on: the whole market moves (term $\beta_i R_M$)
+ ε_i: some randomness due to characteristics proper to the asset (examples of factors specific to the asset: specific announcements such as better-than-expected results for the company, launch of a new product, expected acquisition by another company; or bad news: results worse than expected, social conflicts in the company, new product introduced by a competitor...), independent of the other assets' specific factors (not always satisfied for equities of a same sector, or subsidiaries of a same company...).

Key assumption: correlations across the asset returns only come from the joint effect of the market components of returns. The ε_i are uncorrelated, i.e. there are no effects beyond the market (e.g. industry effects, sector effects, style effects, size effects...) that accounts for the co-movements between security returns.

When the market moves by 1%, the asset moves on average by β%.

$\beta < 1$ ($\beta > 1$) means that the return of an equity or a sector softens (amplifies) the return of the stock market, either downwards or upwards.

Ex: Air Liquide (industrial gases): slope = 0.83, softens the whole market variations.
Carrefour: $R^2 = 0.43$ i.e. 43 % of variations (precisely, return variance) explained by the market returns, 57 % by characteristics proper to the equity.
Note that if R^2 is small, the determination of the regression line is less robust (measure of beta less precise, more sensitive to a given point).
On the main financial markets, on average 40 % of variations are explained by the market returns.

Diversifiable risk and systematic risk

The risk is measured by the variance. The total variance of asset i return is:
$V(R_i) = \beta_i^2 V(R_M) + V(\varepsilon_i)$.

1st term represents the part of the total variance of asset i return which is explained by the variance of the market portfolio (factor R_M common to all assets, softened or amplified by beta).

2nd term = part of the total variance of asset i return which comes from proper characteristics of the asset.

Note that while the statistical explanation of the R^2 is that it provides a measure of the goodness of fit of the regression, the economic rationale is that it provides an estimate of the proportion of the risk of a stock that can be attributed to market risk (as $R^2 = \frac{\beta_i^2 V(R_M)}{V(R_i)}$); the balance $(1 - R^2)$ can then be attributed to firm-specific risk.

For a portfolio X: $R_X = \sum_{i=1}^N x_i R_i$ then $\beta_X = \dfrac{Cov(R_X, R_M)}{V(R_M)} = \dfrac{\sum_{i=1}^N x_i Cov(R_i, R_M)}{V(R_M)} = \sum_{i=1}^N x_i \beta_i$.

The beta of a portfolio is equal to the weighted average of the betas (linearity of the beta).

$R_X = \sum_{i=1}^N x_i \alpha_i + \beta_X R_M + \sum_{i=1}^N x_i \varepsilon_i$ thus $V(R_X) = \beta_X^2 V(R_M) + \sum_{i=1}^N x_i^2 V(\varepsilon_i)$

If we put $\frac{1}{N}$ in each asset, $V(R_X) = \beta_X^2 V(R_M) + \sum_{i=1}^N \frac{1}{N^2} V(\varepsilon_i)$.

For N large, the 2nd term tends to 0 = diversification.

The part of a stock's risk that can be eliminated is called diversifiable risk (or specific risk), while the part that cannot be eliminated is called systematic (= market risk).

$V(\varepsilon_i)$ = non systematic risk or idiosyncratic risk (= specific company risk that can be eliminated through diversification).

For a given asset, from knowing only beta, we know the non-diversifiable risk (and consequently the contribution of the asset to the risk of the portfolio).

On main markets, from $N = 15$ assets,
95 % of the specific risk can be eliminated.

Our 1st version of this property (Markowitz) was using the covariance matrix.
Here (2nd version, from Sharpe), we use the market model which is a simplification (see below).

An **efficient** portfolio X is a combination of the risk-free asset and M, then $R_X = (1-x)r + xR_M$ with x = proportion of wealth invested in portfolio M. Then $\varepsilon = 0$ (no idiosyncratic risk),
$Cov(R_X, R_M) = xV(R_M)$ thus $\beta_X = x$, $\sigma(R_X) = \beta_X \sigma(R_M)$, and $Corr(R_X, R_M) = 1$.

Conclusion: Any efficient portfolio is perfectly diversified; its risk reduces to the systematic risk.

For a well-diversified portfolio, $V(\varepsilon)$ is close to 0, R^2, or $Corr(R_M, R_X)$, is close to 1.

In this model, the covariance matrix can be obtained simply from the betas:
we have, from the assumptions on the model, for any $i \neq j$, $Cov(R_i, R_j) = \beta_i \beta_j V(R_M)$.

It is then much simpler in terms of amount of required input data:
N specific asset's returns components α_i, N betas, N residual risks $\sigma(\varepsilon_i)$, $\mathbb{E}(R_M)$ and $V(R_M)$,

and in terms of computations. In particular Σ^{-1} can be explicitly computed through the following closed formula (see exercise 19):

$$\Sigma^{-1} = D^{-1} - \frac{V(R_M)}{1 + V(R_M)^2\, {}^t\beta D^{-1}\beta} D^{-1}\beta\, {}^t\beta D^{-1} \quad \text{with } \beta = \begin{pmatrix} \beta_1 \\ \dots \\ \beta_N \end{pmatrix} \text{ and } D = diag(V(\varepsilon_1), ..., V(\varepsilon_N)).$$

Note that to invest in M would mean buying some of each asset, and would have a cost (transaction cost). A proxy portfolio is then used. Economic activity corresponds roughly to 15 sectors.
Take one representative asset in each sector (proportion = total capitalisation of the sector).
For CAC40, 20 assets explain most of the diversification.
Some constraints can be introduced, e.g. limits on the x_i (e.g.: none above 15%...).

2.2.4 CONCLUSION OF CAPM

The CAPM gives a **relationship between expected return and non-diversifiable risk** for any traded asset i:

$$\mathbb{E}(R_i) - r = \Pi_M \beta_i \sigma(R_M), \quad \text{while its risk is measured by} \quad V(R_i) = \beta_i^2 V(R_M) + V(\varepsilon_i).$$

The expected premium on an asset depends on its systematic risk, $\beta_i \sigma(R_M)$, not on its total risk.

The investor gets compensated for the systematic risk, not for the diversifiable risk.

ε increases the risk without increasing the expected return.

- To bear a risk, investors require an additional return (compensation for the risk).

- in asset total risk, only non-diversifiable (systematic) risk matters, measured by the beta. The non-diversifiable risk is considered as a drawback (disadvantage), not the specific risk (can be eliminated). The market does not reward investors for bearing diversifiable risk.

- for efficient portfolios, both are equal i.e. total risk is non-diversifiable.

- then the asset's return can be written as r + a risk premium. This risk premium is beta × the risk premium provided by the market ($\mathbb{E}(R_M) - r$). Otherwise stated, at equilibrium, asset risk premium is proportional to Π_M, the average risk aversion of the investors.

- in this model, the expected return of any asset depends only on the beta of this asset.

- this relation risk / expected return holds only for traded assets (it says that any security or portfolio is on the Security Market Line) and results from the market equilibrium: the prices adjust until this relation holds.

For a given asset, if the risk increases (but the average anticipated future value is unchanged), the investors will not accept paying the same price for the asset. Its price goes down, thus its expected return increases. ⇒ the relation is about expected return, but it determines the price today (thus CAPM is a "pricing model").

2.3 CAPM use

Expectation and variance of the returns are estimated from the past, assuming that the probability distribution of the returns is stationary when time passes. Same for the betas: to estimate the beta for the next period, past data are used, thus beta needs to be stable.

Beta estimation

Several companies provide betas estimates: Merrill Lynch, S&P, Morningstar, Bloomberg...) and usually do not reveal their estimation procedure (Bloomberg does).

Data choice: Using a different time period for the regression or different return intervals for the same period can result in a different beta.

⋆ Frequency: weekly, monthly...? Daily is less reliable in particular if trading volumes are low. Betas will be higher for a longer frequency. Generally weekly returns are used.

⋆ Length of the series? Most of the providers use 5-year periods for the estimation (Bloomberg uses 2-year).

⋆ Choice of the reference index: a proxy portfolio is used for computing the market portfolio returns, generally the index of the market in which the stock trades, but for a foreign investor, international indices can be used. Note that while the S&P 500 and the NYSE Composite (representing all common stock listed on the NYSE, including foreign companies) are the most widely used indices for U.S. stocks, they are, at best, imperfect proxies for the market portfolio in the CAPM, which is supposed to include all assets. The S&P 500 represents 80% of the U.S. capitalisation.

⋆ Frequency of the update for the estimates.

⋆ Dividends included or not... The difference can be high for companies paying dividends not in line with the market average (no dividends or high ones).

Additionally, most providers adjust the beta obtained from the usual regression. Bloomberg performs a regression on the S&P500 (for US stocks) using weekly returns over a 2-year period and reports an adjusted beta $= \frac{2}{3}\times$ usual beta $+ \frac{1}{3} \times 1$ (average with the market beta, assuming that a security's beta moves toward the market average over time, i.e. mean reversion, for beta estimated on successive periods). The idea is to obtain an estimate of the security's future beta.

Example: Estimating a Regression Beta for Boeing (source: Aswath Damodaran *Investment Valuation-Tools and Techniques for Determining the Value of any Asset*)

The stock Boeing is traded on the NYSE. We take its monthly returns from January 1996 to December 2000, including both dividends and price appreciation, for the stock and for the market.
The slope of the regression is 0.56. This is Boeing's beta, based on monthly returns from 1996 to 2000 (would be 0.57 if dividends were not taken into account, like Bloomberg does).

Intercept of the regression = 0.54%. This is a measure of Boeing's performance, when it is compared with $r(1-\beta)$. For r, we take the average monthly risk-free rate (since the returns used in the regression are monthly returns) between 1996 and 2000, we get 0.4%.
So the Jensens alpha is $0.54 - 0.4(1 - 0.56) = 0.36\%$.
This analysis suggests that Boeing performed 0.36% better than expected, when expectations are based on the CAPM and on a monthly basis between January 1996 and December 2000. This results in an annualised excess return of approximately $(1 + 0.0036)^{12} - 1 = 4.41\%$.

R^2 of the regression = 9.43%. This statistic suggests that 9.43% of the risk (variance) in Boeing comes from market sources and that the balance of 90.57% of the risk comes from firm-specific components. The latter risk should be diversifiable and therefore will not be rewarded with a higher expected return. Boeings R^2 is higher than the median R^2 of companies listed on the New York Stock Exchange, which was approximately 19% in 2000.

Empirical results: For 2nd approach: CAPM works fine on long periods (i.e. there exists indeed a linear relationship between actual returns and betas), but better for portfolios than for single securities. The betas are more stable for portfolios with a large number of securities than for single securities (ex: JP Morgan between Oct 2015 and Oct 2017 fluctuates between 0.5 and 2).

The "CAPM 1st approach" empirical validation is more complex. First question: is the market portfolio efficient? Also the model assumes that investors make all the same forecasts for mean and variance of returns. In reality, it is considered that views discrepancies generate most transactions on the markets!

Negative or low beta? There are few fundamental investments with consistent and significant negative betas, but some derivatives like equity put options can have large negative betas.
Also: an inverse exchange-traded fund or a short position.
A negative beta might occur even when both the benchmark index and the stock under consideration have positive returns. It is possible that lower positive returns of the index coincide with higher positive returns of the stock, or vice versa. The slope of the regression line in such a case will be negative.

Utility stocks commonly show up as examples of low beta. These have some similarity to bonds, in that they tend to pay consistent dividends, and their prospects are not strongly dependent on economic cycles. They are still stocks, so the market price will be affected by overall stock market trends, even if this does not make sense.

Gold as zero-beta? In their first study, McCown and Zimmerman use the CAPM. Running a regression analysis from 1970 to 2003 where the risk free rate is the yield on the US T-Bill, they use three different proxies for the market portfolio: MSCI US Index, MSCI World denominated in local currency, and MSCI World denominated in US dollar (MSCI, formerly Morgan Stanley Capital International, is a provider of equity, fixed income, hedge fund stock market indexes, some of them calculated since 1969).
Acknowledging that the CAPM is "limited in usefulness as a tool for investment analysis," the authors observe that "all of the estimated beta coefficients for gold are statistically indifferent from zero," and therefore gold does not reflect "much, if any, systematic risk". In this study their conclusion is that gold effectively behaves as a zero-beta asset, yet it also "has a positive risk premium".
Next, McCown and Zimmerman evaluate the investment potential of gold using the Arbitrage pricing theory (APT) of Ross (1976). APT looks at an asset's exposure to multiple sources of risk simultaneously, see below. The study considered various types of market risk including the MSCI World Index denominated in US dollar; bond default spread; bond term spread; change in US industrial production; and change in the log US CPI inflation rate. Regressions were run both with and without the market risk factor.
The result of this analysis shows that gold has positive, statistically significant coefficients on industrial production factors and change in inflation risk factors. As to market risk, for the quarterly and monthly data "the R-squares for gold are all close to zero". McCown and Zimmerman conclude that gold is a good hedge against inflation and a useful addition to portfolios based on APT as a framework.

3 An extension to the mean-variance portfolio analysis: Black-Litterman

Tries to remedy to below limits of the Markowitz optimisation:

- complete set of expected returns is required
- highly-concentrated portfolios obtained (extreme portfolios)
- input-sensitivity (mean-variance portfolios are unstable and often "error maximizers")
- no way to incorporate investor's views.

Bayesian approach to combine the subjective views on expected returns with the market equilibrium vector of expected returns (the prior distribution).
To form a new mixed estimate of expected returns (the posterior distribution).
In addition to the return expectations, the investor must specify the degree of confidence he puts onto the stated views.

4 Multiple-factor models, APT

In the CAPM, the expected return of any security can be explained by the market return, which is the only factor. The beta is indeed the measure of the sensitivity of a security's return to the market return. The relation between the risk premium and the beta is linear.
The CAPM relies on restrictive assumptions about agents' preferences or security returns, and its empirical implications have not been confirmed by data.

Moreover more than one factor is needed to explain asset returns. Other models have been developed with factors replacing the market return, in particular the Arbitrage Pricing Theory **(APT)** (Roll-Ross, 70-80s). See for example Copeland and Weston, chap. 7.

The APT relies on the following assumptions:

1. Asset returns are generated according to a linear factor model.
2. This factor model is the same for all investors, i.e. they have homogenous expectations (as for CAPM).
3. The number of assets is close to infinite (to be able to form arbitrage portfolios with close to zero residual risk, see below).
4. The market is perfect (i.e. perfect competition, no transactions costs [same as CAPM])

Precisely: K risk factors $f_1, ..., ...f_K$, each factor normalised so as to have zero expectation. All the K factors and the risk-free asset are linearly independent.
For a given asset i, the return is governed by factors common to all assets + 1 specific factor:

$$R_i = \mathbb{E}(R_i) + \sum_{k=1}^{K} b_{ik} F_k + \varepsilon_i.$$

b_{ik} denotes the sensitivity of asset i return to factor k, F_k denotes the return of factor k with $E[F_k] = 0$ (explains why first term is $\mathbb{E}(R_i)$), and ε_i denotes the residual return of asset i. It is the specific risk of asset i which is not explained by the factors and satisfies: $E[\varepsilon_i] = 0$

The residuals are non correlated and non correlated with the factors: $Cov[\varepsilon_i, \varepsilon_j] = 0, Cov[\varepsilon_i, F_k] = 0$. Similar to the CAPM, the unique effects are independent and will be diversified away in a large portfolio.

When there exists only one factor corresponding to the market return, this model is the CAPM.

Non-arbitrage conditions lead to the existence of factor risk premia (λ_k) such that:

$$\mathbb{E}(R_i) - r = \sum_{k=1}^{K} b_{ik} \lambda_k \ .$$

For 1 factor, we had $\mathbb{E}(R_i) - r = \beta_i[E(R_M) - r]$ in the first approach.

This can be proved as well in the second approach by arbitrage arguments:

assume $R_i = \alpha_i + \beta_i F + \varepsilon_i$, for $i = 1, 2$, with F centered, then $E(R_i) = \alpha_i$.

1. Assume first that $\varepsilon_i = 0$ for $i = 1, 2$ (no specific risk).

We build a portfolio X such that $R_X = xR_1 + (1 - x)R_2$. This portfolio is risk-free as soon as x is such that $x\beta_1 + (1 - x)\beta_2 = 0$ (canceling the terms in F in R_X).

Then its return can only be r, we get: $x\alpha_1 + (1 - x)\alpha_2 = r$ hence

$x(\alpha_1 - r) + (1 - x)(\alpha_2 - r) = 0$. Multiplying by β_2 and replacing $(1 - x)\beta_2$ by $-x\beta_1$, we obtain: $\beta_2 x(\alpha_1 - r) = x\beta_1(\alpha_2 - r) = 0$ hence $\dfrac{\beta_2}{\beta_2 - \beta_1}(\alpha_1 - r) = \dfrac{\beta_2}{\beta_2 - \beta_1}(\alpha_2 - r)$

We get $\dfrac{\alpha_1 - r}{\beta_1} = \dfrac{\alpha_2 - r}{\beta_2}$ thus $\dfrac{E(R_1) - r}{\beta_1} = \dfrac{E(R_2) - r}{\beta_2} = C^{te}$ (risk-premium for the factor F).

This result generalises into the case with several risky factors, leading to the relationship

$$\mathbb{E}(R_i) - r = \sum_{k=1}^{K} b_{ik} \lambda_k \quad \text{(exercise 20).}$$

2. We ignored the specific risk. One of the main arguments of APT is that with a "large" number of assets the specific risk of a well diversified portfolio is close to zero. To make the argument go through we would need an infinite number of assets. If we have a finite number of assets the relationship above is only an approximate pricing rule. In fact, APT relies on the concept of asymptotic arbitrage. Investors form arbitrage portfolios which contain idiosyncratic risk, then when we take limits of these portfolios (as the number of assets go to infinity) they become free of residual risk (as in above argument).

2 types of multi-factor models:

Exogenous models (macroeconomic and fundamental models). The common factors are predetermined factors such as explicit macroeconomic variables as market return, gross domestic product growth, long-short interest rate spread, credit spread, inflation rate, exchange rate,... , and microeconomic variables which are determined cross-sectionally from normalized fundamental attributes (price earning ratios, earning yield, earning variability, currency sensitivity, dividend yield, book-to-market ratio, size, industry, leverage, trading activity, momentum...).

Endogenous models (statistical models). The risk factors are extracted directly from the return data by using statistical methods such as factor analysis, Principal Component Analysis (PCA)... then a factor can be interpreted economically by identifying an economic variable with which it is strongly correlated (not very satisfying though).

Numerous empirical studies are devoted to the determination of the macroeconomic or financial factors. One problem is also to identify the number of factors.

Summary: Arbitrage pricing theory is a general theory of asset pricing that states that the future return of a financial asset can be modeled as a linear function of various macro-economic factors or theoretical market indices, where sensitivity to changes in each factor is represented by a factor-specific beta coefficient. The model-derived rate of return will then be used to price the asset correctly - the asset price should equal the expected end of period price discounted at the rate implied by the model. If the price diverges, arbitrage should bring it back into line.

CAPM based on the conjecture of an equilibrium market.
APT based on the conjecture of the absence of arbitrage.

The APT offers an alternative to the CAPM. APT does not require assumptions about utility or the distribution of security returns (but the risk is still measured by the return variance). Factor models are often used in practice to reduce the dimensionality of the estimation problem in the Markowitz approach, like it was already the case with the one factor CAPM (the matrix inversion is simpler as well). K factors imply more computation than 1 factor, but still less than computing the covariance of each possible pair of securities in the portfolio.

APT is a generalisation of CAPM (2nd approach), but CAPM remains one of the favorite tools of practitioners, simple, and there exists already a well-established tradition of operating betas.

For example, the three-factor model of Fama and French (1993) takes account of the book-to-market ratio (ratio of the book value of a firm to its market value; allows to identified under-valued stocks) and the company's size measured by its market capitalization:

$$R_i - r = \alpha_i + \beta_i(R_M - r) + b_i^S SMB + b_i^H HML + \varepsilon_i$$

where SMB indicates small (cap) minus big. It denotes the difference between returns of the small-capitalization stocks and large-capitalization returns.
The term HML is the high (book/price) minus low. It denotes the difference between returns on portfolios with high book-to-market ratios and portfolios with low book-to-market ratios (or excess returns of value stocks over growth stocks).

On a given example, the Fama and French three-factor model explains over 90% of the diversified portfolios returns, compared with the average 70% given by the CAPM.

In 2015, Fama and French extended the model, adding a further two factors - profitability and investment. Defined analogously to the HML factor, the profitability factor (RMW) is the difference between the returns of firms with robust (high) and weak (low) operating profitability; and the investment factor (CMA) is the difference between the returns of firms that invest conservatively and firms that invest aggressively.

Chen, Roll, Ross (86) have identified the following principal factors on the American equities market:

Growth rate in industrial production,
Expected inflation,
Unexpected inflation,
Changes in risk premiums,
Changes in the term structure (yield curve),
Return on the market portfolio.

Finally, note that while multi-factor models have been originally developed for simplifying the optimization process, they nowadays are mainly used for hedging in the management of passive/active portfolios (tracking portfolios, factor bets, performance and risk analyses).

Chapter III Performance, dependence and risk measures

J.L. Prigent (2007)
Portfolio Optimization and Performance Analysis, Chapman & Hall,

P. Embrechts, A. McNeil,. D. Straumann (1999)
Correlation and Dependence in Risk Management: Properties and Pitfalls.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D. (1999)
Coherent measures of risk. Math. Fin. 9(3).

1 Some performance measures

Three standard performance measures based on the Capital market line and the Security market line. The higher these performance measures, the more interesting the portfolio.

1.1 Sharpe ratio

This measure is based on the Capital market line. For all efficient portfolios, we have the following equality between (excess reward)/(total risk) ratios:

$$\frac{\mathbb{E}\,(R_P) - r}{\sigma(R_P)} = \frac{E(R_M) - r}{\sigma(R_M)}.$$

The previous ratio is the slope of the capital market line. As proposed by Sharpe (1966), this can be actually considered as a performance measure. It is equal to the excess mean return and the measure ot total risk (the standard deviation), and is defined as the reward-to-variability ratio. For example, the manager can check if the excess mean return of the portfolio is sufficient to compensate a higher risk than the market portfolio. If the portfolio is well-diversified, its Sharpe ratio is close to the market portfolio's.

1.2 Treynor measure

The Treynor's ratio (1966) is directly based on the CAPM: $\frac{\mathbb{E}\,(R_P) - r}{\beta_P}$. It can also be viewed as a reward-to-risk ratio where the risk is the exposition to market risk. At equilibrium, this ratio is constant and equal to $\mathbb{E}\,(R_M) - r$. The Treynor ratio allows us to evaluate the performance of a well diversified portfolio, since it only involves the systematic risk. It can be used to examine performance of portfolio which is only a part of the investor's assets. Due to previous diversification, the investor takes care only of the systematic risk.

1.3 Jensen's alpha

This measure is also based on the CAPM, but mainly when it is not satisfied. Indeed, it is computed as the difference α_P between the mean return of portfolio P and the return explained by the CAPM, $r + \beta_P(\mathbb{E}\,(R_M) - r)$, i.e. the Jensens alpha is defined as:

$$\alpha_P = \mathbb{E}\,(R_P) - r - \beta_P(\mathbb{E}\,(R_M) - r)$$

The manager searches for portfolios with $\alpha_P > 0$. For ex-post measures, the manager will be judged skillful if the coefficient α_P is significantly above 0.

If $\alpha_P = 0$, the return of portfolio P is at equilibrium and the manager's forecasts have not beat the market performance: the portfolio has the same alpha as any combination of the risk-free asset and the market portfolio. Unlike the Sharpe and Treynor ratios, the Jensen measure contains the benchmark itself. As with the Treynor ratio, it only takes account of the systematic risk.

The CAPM states that the expected value of alpha is zero for all securities.

Due to the particular form of alpha, only portfolios with the same risk beta can be compared. Otherwise, we can consider, for example, the Black-Treynor ratio: $\frac{\alpha_P}{\beta_P}$.

1.4 Sortino ratio

Indicators based on standard deviations like the Sharpe ratio do not indicate whether the value is above or below the mean.

Risk-adjusted measures based, for example, on semi-variance (defined as $\mathbb{E}\left(\{([R_P - \mathbb{E}(R_P)]^+)^2\}\right)$ can separate these two cases. Thus, they can better take account of asymmetrical return distributions. This is the purpose of the Sortino ratio, introduced in Sortino and Price (1994).

It is defined similar to the Sharpe ratio, but with a minimum acceptable return (MAR) instead of the risk-free return. Moreover, the standard deviation is replaced by the semi-standard deviation of the return below the MAR:

$$\frac{\mathbb{E}(R_P) - MAR}{\sqrt{\mathbb{E}\{[(MAR - R_P)^+]^2\}}}$$

2 How to measure the dependence

2.1 Correlation

The most commonly used dependence measure is Pearson's linear correlation coefficient (for pairs of real-valued, non degenerate random variables X, Y with finite variances). The linear correlation coefficient $Corr(X,Y)$ is good at characterizing linear relationships between X and Y : $|Corr(X,Y)| = 1$ for perfectly deterministic linear relationship ($Y = aX + b$ a.s. or $P[Y = aX + b] = 1$ for $a \in \mathbb{R}\ 0$, $b \in \mathbb{R}$), and in the case of independent random variables, $Corr(X,Y) = 0$. However, it does not measure the nonlinear relationships between X and Y well.

The correlation coefficient is strictly related to the slope parameter of a linear regression of the r.v. Y on the r.v. X, and it measures only the co-dependence between the linear components of X and Y. Indeed, we saw:

$$Corr(X,Y)^2 = \frac{V(Y) - min_{a,b} V(Y - (a + bX))}{V(Y)}.$$

The regression coefficients a, b are given by $b = \frac{Cov(X,Y)}{V(X)}$ and $a = \mathbb{E}(Y) - b\mathbb{E}(X)$.

For this (a,b), $\frac{V(Y) - V(a + bX)}{V(Y)} = \frac{V(\varepsilon)}{V(Y)}$ (where ε is the residual $Y - (a + bX)$) is the relative variation of $V(Y)$ by linear regression on X.

It is small when X and Y are very correlated ($Corr(X,Y)$ close to -1 or 1).

Correlation fulfills the linearity property $Corr(\alpha X + \beta, \gamma Y + \delta) = sgn(\alpha\gamma) Corr(X,Y)$, when

$\alpha, \gamma \in \mathbb{R}^*$, $\beta, \delta \in \mathbb{R}$. Correlation is thus invariant under positive affine transformations, i.e. strictly increasing linear transformations.

The variance of any linear combination is fully determined by the pairwise covariances between the components (we found $V(X) = {}^tX\Sigma X$).

+ naturalness as a measure of dependence in multivariate normal distributions and, more generally, in multivariate spherical and elliptical distributions (see below).

2.2 Spherical and elliptical distributions

Reminder: A random vector X is a Gaussian vector (or has a Gaussian distribution) iff its characteristic function has the form:

$$\forall \lambda \in \mathbb{R},\ \mathbb{E}\left(e^{i\,{}^t\lambda X}\right) = e^{i\,{}^t\lambda\mu - \frac{1}{2}\,{}^t\lambda\Sigma\lambda}$$

for some symmetric positive semi-definite $n \times n$ matrix Σ and some vector $\mu \in \mathbb{R}^n$.

Then μ is the mean vector of X i.e. ${}^t\mu = (\mathbb{E}\,(X_1), ..., \mathbb{E}\,(X_n))$ and Σ is the covariance matrix of X i.e. $Cov(X_j, X_k) = \mathbb{E}\,[(X_j - \mu_j)(X_k - \mu_k)] = \Sigma_{jk}$. Thus, a Gaussian vector is completely characterized by its mean vector and covariance matrix. Its component are independent iff its covariance matrix is diagonal.

Th: If X is a Gaussian vector with mean μ, there exists independent Gaussian variables $(Y_1, ..., Y_n)$, centered, and an orthogonal matrix A s.t. $X = AY + \mu$.
Reminder: a matrix A is orthogonal iff it satisfies $A\,{}^tA = {}^tAA = I$, i.e. its columns and rows are orthonormal vectors.

Proof: Σ is a symmetric matrix hence it is diagonalisable by an orthogonal matrix:
There exists A orthogonal and Λ diagonal such that $\Sigma = A\Lambda\,{}^tA$. Note that ${}^tA\Sigma A = \Lambda$.
As Σ is semi-positive, the coefficients of Λ are non negative.
Let $Y = {}^tA(X - \mu)$. It is a centered vector (each element is the linear combinaison of some centered variables). We have for any $\lambda \in \mathbb{R}$,
$\mathbb{E}\left(e^{i\,{}^t\lambda Y}\right) = \mathbb{E}\left(e^{i\,{}^t\lambda\,{}^tA(X-\mu)}\right) = e^{-i\,{}^t(A\lambda)\mu}\,\mathbb{E}\left(e^{i\,{}^t(A\lambda)X}\right) = e^{-\frac{1}{2}\,{}^t\lambda\,{}^tA\Sigma A\lambda} = e^{-\frac{1}{2}\,{}^t\lambda\Lambda\lambda}$.
Then Y is a Gaussian vector with a covariance matrix equal to Λ.

The distribution of a Gaussian vector X has a density iff Σ is positive definite, then, the density is:

$$\forall (x_1, ..., x_n) \in \mathbb{R}^n, \quad f_X(x_1, ..., x_n) = \frac{1}{\sqrt{(2\pi)^n det\Sigma}} e^{-\frac{1}{2}\,{}^t(x-\mu)\Sigma(x-\mu)}$$

A *sphere* in dimension n is a set of vectors X such that $\sum_{i=1}^{n} x_i = R^2$ i.e. ${}^tXX = R^2$.

An *ellipsoid* is a surface that may be obtained from a sphere by deforming it by means of directional scalings, or more generally, of an affine transformation. An arbitrarily oriented ellipsoid, centered at μ, is defined by the solutions X to the equation $({}^tX - \mu)\Sigma\,(X - \mu) = 1$, where Σ is a positive definite matrix and X, μ are vectors. Ellipsoids are transforms of spheres by maps $X \mapsto AX + \mu$.

The spherical distributions extend the standard multivariate normal distribution $\mathcal{N}(0, I)$, i.e. the distribution of independent standard normal variables. They provide a family of symmetric distributions for uncorrelated random vectors with mean zero.

Def: A random vector $X = {}^t(X_1, ..., X_n)$ has a spherical distribution if for every orthogonal matrix $n \times n$ A, AX and X have same distribution.

Such distributions have characteristic functions with a particularly simple form:
$\mathbb{E}\,[e^{i\,{}^t\lambda X}] = \Phi({}^t\lambda\lambda) = \Phi(\lambda_1^2 + ... + \lambda_n^2)$.
The function Φ is called the characteristic generator of the spherical distribution.

If X has a density (probability density function, or pdf) $f(x) = f(x_1, ..., x_n)$ then this is equivalent to $f(x) = g({}^txx) = g(x_1^2 + ... + x_n^2)$ for some function $g : \mathbb{R}^+ \longleftarrow \mathbb{R}^+$, so that the spherical distributions are best interpreted as those distributions whose density is constant on spheres.

Note that these are the distributions of uncorrelated random variables but, contrary to the normal case, not the distributions of independent random variables. In the class of spherical distributions the multivariate normal is the only distribution of independent random variables (see Symmetric multivariate and related distributions. Fang, Kotz, Ng, Chapman and Hall,1987).

Elliptical distributions are the affine maps of spherical distributions in $\mathbb{R}^n$, i.e. $AX + \mu$ for X having a spherical distribution, and with $A \in \mathbb{R}^{n\times n}$, $\mu \in \mathbb{R}^n$.

Such distributions have for characteristic functions:
$\forall\lambda \in \mathbb{R}^n,\ \mathbb{E}\,[e^{i\,{}^t\lambda(AX+\mu)}] = e^{i\,{}^t\lambda\mu}e^{i{}^t({}^tA\lambda)X} = e^{i\,{}^t\lambda\mu}\Phi({}^t\lambda A\,{}^tA\lambda)) = e^{i\,{}^t\lambda\mu}\Phi({}^t\lambda\Sigma\lambda)$, where $\Sigma = A\,{}^tA$.
Denoted by $X \sim E_n(\mu, \Sigma, \Phi)$. Again, Φ is called the characteristic generator of the distribution.

For example, multivariate Gaussian distribution $N_n(\mu, \Sigma,) = E_n(\mu, \Sigma, \Phi)$ with $\Phi(\lambda) = e^{-\frac{\lambda^2}{2}}$, multivariate Student and logistic distributions...

If X has a density $f(x) = g({}^txx) = g(x_1^2 + ... + x_n^2)$ and if A is invertible (then Σ is positive definite), $AY + \mu$ has density $h(x) = \dfrac{1}{\sqrt{det(\Sigma)}} g({}^t(x - \mu)\Sigma^{-1}(x - \mu))$.

Elliptical distributions are those distributions whose density is constant on ellipsoids.

Many of the properties of the multivariate normal distribution are shared by the elliptical distributions. Linear combinations, marginal distributions and conditional distributions of elliptical random variables can largely be determined by linear algebra using knowledge of covariance matrix, mean and generator:

- any linear combination of an elliptically distributed random vector is also elliptical with the same characteristic generator Φ.
Proof: If $X \sim E_n(\mu, \Sigma, \Phi)$ and $B \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, then $BX + b \sim E_m(B\mu + b, B\Sigma\,{}^tB, \Phi)$.

- the marginal distributions of elliptical distributions are also elliptical with the same generator.
Proof: compute the characteristic function at $\lambda = (0, ..., 0, \lambda_k, 0, ..., 0)$.

2.3 Shortcomings of correlation

But the correlation is not sufficient to describe the dependence for most financial returns and has several drawbacks. For example:

- The variances of X and Y must be finite else the linear correlation is not defined.

- Zero correlation does not in general imply independence.

- While correlation is invariant under strictly increasing linear transformations, it is not invariant by nonlinear such transformations, i.e. we generally have $Corr(T(X), T(Y)) \neq Corr(X, Y)$ for $T : \mathbb{R} \to \mathbb{R}$ nonlinear strictly increasing transformation.
Note that for bivariate normally-distributed vectors, one can show that for arbitrary transformations $T : \mathbb{R} \to \mathbb{R}$, $|Corr(T(X), T(Y))| \leq |Corr(X, Y)|$).

Fallacy 1. Marginal distributions and correlation **do not** determine the joint distribution.
The distribution of (X, Y) is not uniquely determined by F_X, F_Y and $Corr(X, Y)$!
This is true if we restrict our attention to the multivariate normal distribution or the elliptical distributions. For example, if we know that (X, Y) have a bivariate normal distribution, then the expectations and variances of X and Y and their correlation uniquely determine the joint distribution. However, if we only know the marginal distributions of X and Y and the correlation then there are many possible bivariate distributions for (X, Y).

The class of r.v. for which linear correlation can be used as a dependence measure is the class of elliptical distributions. In particular, the mean-variance (Markowitz) approach is suited to (and only to) the case of elliptical distributions like normal or t-Student distributions with finite variances.

Attainable correlation

Fallacy 2. Given marginal distributions F_X and F_Y for X and Y, when changing their joint distribution, all linear correlations between -1 and 1 **cannot** be attained.

Linear correlation attains the bounds ± 1 only if X and Y are non-trivial affine transformations of one another, i.e.: $r(X, Y) = 1 \Longleftrightarrow Y = a + bX,\ b > 0$
$r(X, Y) = -1 \Longleftrightarrow Y = a + bX,\ b < 0.$

Th (admitted): For any variables (X, Y) with given marginal laws such that $0 < V(X), V(Y) < +\infty$, the **possible** values for the **correlation** $Corr(X, Y)$ is a closed interval $[\rho_{min}, \rho_{max}]$ containing 0. $|\rho_{min}|$ and $|\rho_{max}|$ may be **dramatically** smaller than 1:

$$\rho_{min} \leq Corr(X, Y) \leq \rho_{max}, \text{ with } \rho_{min} < 0 < \rho_{max}.$$

Proof: uses let F_X and F_Y marginal distributions of X and Y, then $Cov(X, Y) = \int\int [F(x, y) - F_X(x)F_Y(y)]dxdy$ (Hoeffding) and Fréchet bounds for copula: $\max(x_1 + \ + x_n + 1 - n, 0) \leq C(x_1, ..., x_n) \leq \min\{x_1, ..., x_n\}$.

Comonotonicity and countermonotonicity
The random variables $X_1, ..., X_n$ are said to be comonotonic if $(X_1, ..., X_n) = (h_1(Z), ..., h_n(Z))$ in distribution, where Z is some random variable, and $h_1, ..., h_n$ are increasing functions. ˇ

The random variables X_1, X_2 are said to be countermonotonic if $(X_1, X_2) = (h_1(Z), h_2(Z))$ in distribution, where Z is some random variable, and h_1 and h_2 are respectively an increasing and a decreasing function, or vice versa.

Let X and Y be two random variables with strictly positive and finite variances, and let ρ_{min}

and ρ_{max} denote the minimum and maximum possible correlation coefficient between X and Y. Then, it can be shown that:

Th (admitted): $Corr(X, Y) = \rho_{min}$ if and only if X and Y are countermonotonic;
$Corr(X, Y) = \rho_{max}$ if and only if X and Y are comonotonic. ˇ

Example of $[\rho_{min}, \rho_{max}]$ dramatically different of $[-1, 1]$: attainable correlation for lognormal random variables X and Y s.t. $\ln X \sim \mathcal{N}(0, 1)$ and $\ln Y \sim \mathcal{N}(0, \sigma^2)$.

We take X and Y comonotonic to obtain ρ_{max}, see exercise 22.

$\lim_{\sigma \to -\infty} \rho_{min} = \lim_{\sigma \to +\infty} \rho_{max} = 0$, so the possible correlations between 2 lognormal variables s.t. $\ln X \sim \mathcal{N}(0, 1)$ and $\ln Y \sim \mathcal{N}(0, \sigma^2)$ are very limited for σ large.

The linear correlation coefficient can be almost zero, even if X and Y are comonotonic or countermonotonic!

2.4 Copula

Let X be a random variable with cumulative distribution function (cdf) F_X.
A cdf takes values in $[0, 1]$, it is always increasing and continuous from the right but it can be constant on some intervals or have jumps, then it is not always invertible.

The generalised inverse distribution function is defined by:
for $p \in [0, 1]$, $F_X^{-1}(p) = \inf\{x | F_X(x) \geq p\}$ (also called the quantile function).

Prop: **1.** For any r.v. $U \sim U(0, 1)$ (uniformly distributed on $[0, 1]$), $F_X^{-1}(U)$ has same law as X. (note that this gives a simple method for simulating random variates with cdf F_X.)

2. If F_X is continuous, then $F_X(X) \sim U(0, 1)$.

Proof: For sake of simplicity, we assume that F_X is continuous (i.e. X is a continuous random variable), i.e. $P(X = x) = 0$ for all $x \in \mathbb{R}$, and strictly increasing, then F_X^{-1} is just its inverse function. Both results are straightforward in that case:

1. let $x \in \mathbb{R}$, $P(F_X^{-1}(U) \leq x) = P(U \leq F_X(x)) = F_X(x)$ (as F_X is increasing).

2. let $p \in]0, 1[$, $P(F_X(X) \leq p) = P(X \leq F_X^{-1}(p)) = F_X(F_X^{-1}(p)) = p$.

We consider now n real-valued random variables. In fact, to describe completely the dependence between the real-valued random variables $X_1, ..., X_n$ their joint cumulative distribution function should be used: $F(x_1, ..., x_n) = P[X_1 \leq x_1, ..., X_n \leq x_n]$.

The idea of separating F into a part which describes the dependence structure and parts which describe the marginal behaviour only, has led to the concept of **copula**.

For simplicity we assume that each X_i has a continuous marginal cdf F_i, then $F_i(X_i) \sim U(0, 1)$.

The joint cdf C of ${}^t(F_1(X_1), ..., F_n(X_n))$ is then called the copula of the random vector ${}^t(X_1, ..., X_n)$. We have, for all $(x_1, ..., x_n) \in \mathbb{R}^n$:

$$\begin{aligned} F(x_1, ..., x_n) &= P(X_1 \leq x_1, ..., X_n \leq x_n) = P(F_1(X_1) \leq F_1(x_1), ..., F_n(X_n) \leq F_n(x_n)) \\ &= C(F_1(x_1), ..., F_n(x_n)). \end{aligned}$$

C corresponds to the **dependence structure** of $X_1, ..., X_n$.

This representation allows to take into account the problems connected with dependent extreme events that cannot be characterized via linear correlation, which is very important in finance.

Examples: for independent variables, $C(x_1, ..., x_n) = x_1 \cdot ... \cdot x_n$.

One attractive feature of the copula representation of dependence is that it is invariant under increasing and continuous transformations of the marginals.

Th: If ${}^t(X_1, ..., X_n)$ has copula C and $T_1, ..., T_n$ are increasing continuous functions, then ${}^t(T_1(X_1), ..., T_n(X_n))$ also has copula C.

Proof. Let ${}^t(U_1, ..., U_n)$ have distribution function C (in the case of continuous marginals F_{X_i} take $U_i = F_{X_i}(X_i)$). We have:

$$C(F_{T_1(X_1)}(x_1), ..., F_{T_n(X_n)}(x_n)) = P(U_1 \leq F_{T_1(X_1)}(x_1), ..., U_n \leq F_{T_n(X_n)}(x_n)]$$
$$= P(F^{-1}_{T_1(X_1)}(U_1) \leq x_1, ..., F^{-1}_{T_n(X_n)}(U_n) \leq x_n) = P(T_1 \circ F^{-1}_{X_1}(U_1) \leq x_1, ..., T_n \circ F^{-1}_{X_n}(U_n) \leq x_n)$$
$$= P(T_1(X_1) \leq x_1, ..., T_n(X_n) \leq x_n).$$

Note that for 2 r.v.:

- comonotonicity corresponds to the upper bound possible for any copula (perfect positive dependence), which is the cdf of the random vectors ${}^t(U, U)$ where $U \sim U(0, 1)$.
- countermonotonicity corresponds to the lower bound possible for any copula (perfect negative dependence), which is the cdf of the random vectors ${}^t(U, 1 - U)$.

Indeed:

Equivalent condition for n comonotonic variables: their copula is the Fréchet upper bound $M(u_1, ..., u_n) = min(u_1, ..., u_n)$, which is the strongest type of "positive" dependence. ˇ

Equivalent condition for 2 countermonotonic variables: their copula is the Fréchet lower bound $W(u_1, u_2) = max(0, u_1 + u_2 - 1)$, which is the strongest type of "negative" dependence in the bivariate case. Countermonotonocity doesn't generalize to higher dimensions.

3 Risk measures

3.1 Classical examples beyond variance (or volatility)

Non symmetric distributions in finance may encourage the use of semi-variance, but not really satisfying (main desirable properties for a risk measure missing).

3.1.1 Value-at-risk (VaR)

Concept introduced to answer the following question: how much one can expect to lose in one day, week, year... with a given probability. What is the percentage of the investment that is at risk? What is the minimum loss incurred in the α% worst cases of our portfolio?

X = random variable describing the future value of the P&L of a portfolio (or its return) on some given time horizon T from today.
The value-at-risk at level α, VaR_α, is defined by $P(X \leq -VaR_\alpha) = 1 - \alpha$.

For $\alpha = 95\%$, we are looking at the 5% worst losses.

Let F_X is the cumulative distribution function of X, then $VaR_\alpha = -F_X^{-1}(1-\alpha)$. VaR_α is then the "$(1-\alpha)$-quantile" of the loss distribution.

At a firm level, we can think of the VaR as the amount of extra-capital that a firm needs in order to reduce the probability of going bankrupt to α. This amount should be kept negative.

Note that the definition has to be more precise if F_X is not invertible:
VaR_α is defined as $= \inf\{x|P[X \leq x] > 1-\alpha\}$.

Use imposed in regulation of financial institutions. In particular the major principles of the 2001 proposal of the Basel Banking Supervisory Committee are:
- VaR is assumed as risk measure,
- the risk of each loan must be portfolio invariant, i. e. must be measured by its own characteristic only, not taking into account those of portfolio in which the loan is,
- the regulatory capital for a loan must be correlated to its marginal contribution to VaR.

Computation methods:
- historical (empirical quantile),
- Parametric (based on distributional assumptions),
  for example normal VaR is built under assumption of normality of returns),
  for Risk metrics VaR (a standard for practitioners), an exponential moving-average is used to forecast the volatility, and to compute the quantile assuming a Gaussian distribution.
- the Monte Carlo method estimates VaR by simulating random scenarios and revaluing portfolio positions.

The use of VaR has been generalized, VaR is the most widely used instrument to control risk, but it has several drawbacks:

VaR does not measure loss exceeding VaR;
it may provide conflicting results at different confidence levels;
VaR has many local extremes leading to unstable VaR ranking.

We will see below that some good properties for a risk measure are missing, namely sub-additivity and convexity.

3.1.2 Expected shortfall

Risk measure that is more sensitive than VaR to the shape of the loss distribution in the tail of the distribution.
The expected shortfall measures the expected losses exceeding VaR:
$ES_\alpha = -\mathbb{E}\left[X|X < -VaR_\alpha\right]$ "expected shortfall at $\alpha\%$ level"
= expected return on the portfolio in the worst $\alpha\%$ of the cases.

Close concepts: conditional value at risk (CVaR) and expected tail loss (ETL). All match when the cumulative distribution function of the variable is continuous.

3.1.3 The maximum drawdown

The drawdown is the measure of the decline from a historical peak in some price process (X_t) (typically the Net Asset Value of a financial trading strategy or of a fund).

With $X_0 = 0$, the drawdown at time T is computed as $\left[\max_{t\in(0,T)} X_t - X_T\right]^+$

The maximum drawdown up to time T is the maximum of the Drawdown over the history of the variable:

$$\mathrm{MDD}(T) = \max_{\tau\in(0,T)} \left[\max_{t\in(0,\tau)} X_t - X_\tau\right]^+$$

It is the worst peak to valley loss since the investments inception. This is then an indicator of risk, used particularly in asset management.

3.2 Coherent risk measures

To measure risk is equivalent to establishing a correspondence ρ between the space X of random variables (for instance the returns of a portfolio) and non-negative real numbers, i.e. $\rho : X \mapsto \mathbb{R}^+$ (scalar measures of risk).

A risk measure ρ is **coherent** if it satisfies the following properties:

– translation invariance: $\forall X, \forall m \in \mathbb{R}$, $\rho(X + m) = \rho(X) - m$.

– sub-additivity: $\forall X, Y$, $\rho(X + Y) \leq \rho(X) + \rho(Y)$

– positive homogeneity: $\forall \lambda \geq 0$, $\forall X$, $\rho(\lambda X) = \lambda\rho(X)$

– monotonicity: $\forall X, Y \in \mathcal{X}$ with $X \leq Y$, $\rho(Y) \leq \rho(X)$.

Comments:

- $\rho(X + Y) > \rho(X) + \rho(Y)$ would imply, for instance, that in order to decrease risk, it could be convenient to split up a company into different distinct divisions. From the regulatory point of view this would allow to reduce capital requirements. Note that covariance is subadditive, and this property turned out to be essential in Markowitz portfolio theory.

- Monotonicity rules out any semi-variance type of risk measure.

- Sub-additivity is the mathematical equivalent of the diversification effect.

VaR is not sub-additive, except in the case in which the joint distribution of returns is elliptic. That implies that portfolio diversification may lead to an increase of risk and prevents to add up the VaR of different risk sources.

Counter-example: consider the case of a bank which has given a \$100 loan to a client whose default probability is equal to 0.008. With $\alpha = 99\%$, $VaR_\alpha \leq 0$.

Consider now another bank which has given two loans of \$50 each and for both, the default probability is equal to 0.008. In case the default probabilities of the two loans are independent, VaR_α is \$50.

Hence we have that diversification, which is commonly considered as a way to reduce risk, can lead to an increase of VaR.

Also, VaR is non-convex (weaker than coherent), which makes it impossible to use VaR to measure the risk in portfolio optimization problems.

In the elliptical world, VaR and the expected shortfall are coherent risk measures, the Markowitz variance-minimizing portfolio minimizes both of them. See for example Gupta, Arjun K.; Varga, Tamas; Bodnar, Taras (2013). Elliptically contoured models in statistics and portfolio theory.