$$\delta^2 \approx \frac{n}{n-1} x_{top} \Phi(x_{top})^2 \left(1 + \varepsilon x_{top} \frac{x_{top}}{n-1}\right)\left(x_{top} + \frac{n-1}{n x_{top}} + \varepsilon x_{top}\left(1 + \frac{1}{n} - (1+\varepsilon) x_{top}^3 \frac{\Phi(x_{top})}{n-1}\right)\right)$$

$$-2\delta x_{top}\varepsilon \approx T_{op} = \frac{n}{n-1} x_{top}\left(\Phi(x_{top})\right)^n$$

$$\frac{x_{top}\Phi(x_{top})}{n-1} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}{x_{top}}^2}$$

We first infer that:

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}{x_\delta}^2} \approx \frac{x_{top}\Phi(x_{top})}{n-1}\left(1 - \varepsilon x_{top} x_{top}\right) \approx \frac{x_{top}\Phi(x_{top})}{n-1}$$

Next we would infer that:

$$\delta^2 \approx \frac{n}{n-1} x_{top}\Phi(x_{top})^2 \left(x_{top} + \varepsilon x_{top}\left(1 - x_{top}^3 \frac{\Phi(x_{top})}{n-1}\right)\right) \approx \frac{n}{n-1} x_{top}^2 \Phi(x_{top})^2$$

Thus a first gross estimate would yield: $\frac{\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}{x_\delta}^2}\right)^2}{\delta^2} \approx \frac{x_{top}^2 \Phi(x_{top})^2}{(n-1)^2 \frac{n}{n-1} x_{top}^2 \Phi(x_{top})^2} \approx \frac{1}{n(n-1)} < \frac{1}{30} \ll 1$

This is the square so the approximation runs in $\frac{1}{n-1}$ at worst.

Conclusion: we can estimate well the expectation of the max for n>6 for a normal distribution. The fundamental equations to produce this estimates derive from the expression of the density of the Max that is always the same . Thus we can use on the follow the Hermite's function to proxy the real life distribution. The play of the consecutive power laws will make the formulas more intricate but the principles above remain.

So in the larger scheme, a maximum expected return being indexed by a "risk", this "risk" can be equivalently perceived as another "maximum expected return" that is obtained relative to some inner "risk" and so on and so forth. If therefore the original "risk" has actually a "distribution" because it matches with a distribution of "maximum expected returns" that has some expectation as such, we can refine an initial efficient frontier by defining a maximum expected return for a given bucket of the original distribution of the risk that we initially only defined through the expectation of this distribution. This is how we should add higher moments since they indirectly feature the permanent uncertainty that surrounds our measure of the "risk". We initially assume that the variance provides a good estimate of this risk. But it truly is random, even for the normal distribution. Yet we really care here about the deviation from the normal distribution: indeed the inclusion of the higher moments can make sense only if the CLT is not good enough to capture the "tail risk", the "skew risk" or some other non normal ass

5-Efficient Frontier

The starting point is the Markowitz efficient frontier: one would optimize the weights in a portfolio of assets based on the covariance and the expected returns of the individual assets one relative to the

others. Kristiaan Kerstens, for example, added elements that are based upon the higher statistical moments of the distribution of returns and a wider range tof criteria to measure the many possible investment strategies in order to pick “the optimal one”.

The Markowitz efficient frontier in a nutshell…

We consider a “portfolio” of assets or, more generally, a universe of possible asset classes that we can define through an expected return, a variance and more generally through a distribution of returns.

Let’s consider n assets x_k and a weighted combination made of weights w_k that would exhibit a return $r_w = \sum_{k=1}^{k-n} w_k r_k$

and a variance that would involve a correlation matrix, alternatively a covariance matrix (real symmetrical hence one that can be diagonal but not necessarily having all positive eigen values)

$$v_w = \sum_{k,l=1}^{k,l=n} w_k w_l \rho_{kl} \sqrt{v_k v_l} \ \ where \ \sqrt{v_k} = \sqrt{\frac{1}{N-1}\sum_{j=1}^{j=N}\left(r_{kj} - r_k\right)^2} \ \ with \ r_k = \frac{1}{N}\sum_{j=1}^{j=N} r_{kj}$$

in the expressions above we assume that we dispose of “N” empirical evidence of varying returns as measured through historical data or else Monte Carlo simulations. The key point being that “returns” follow a certain distribution that we have access to in part at least.

Depending upon the range of expected return, depending on the covariance, a weighted combination of the assets (assuming no leverage ie, long risk exposures only in general) will exhibit itself a distribution of returns and a range of expected returns. The “efficient frontier” concept implements the view that for every targeted “expected return” there is an optimal set of weights that will exhibit the lowest possible variance.

We should now express the characteristic of the portfolio in vectors and matrices since it will prove quite useful next.

We thus have a portfolio identified with a vector of weights “w” that would have an expected return based on a vector of individual expected returns (asset per asset) and we would look for the optimal vector of weights that would minimize the variance of the portfolio for a given expected return:

$$return \ r_w = w^T r, variance \ v_w = w^T C w \ C: covariance \ matrix$$

So if we write this into an “optimization” problem we have:

$$\text{Minimize } v_w = w^T C w \ under \ the \ constraints: \begin{cases} w^T r = r_{fixed} \\ w_i \geq 0 \\ w_1 + w_2 + w_3 + \cdots w_n = 1 \end{cases}$$

This problem is typical of a simplex problem that has a solution in general. We find in general a “risk aversion” pattern in historical data where the higher the target return we have the higher the increase in risk terms is going to be. In other words the efficient frontier is concave: if we target a 1% return we would be able to face say a 2% risk. If we target next a 2% return we would likely have to face a 5% risk instead of 2*2%=4%. And if we target a 4% risk, we likely will face a 12.5% risk instead of 10% ie 2*5% or even less so 8% as 4*2%.

When there is the possibility to generate financial leverage for some assets, then some players ,having more risk appetite because they can build an arbitrage on this risk aversion pattern by placing “long short” exposures through higher trading and risk management skills, will indeed put on

"trades" that will alter the shape of the efficient frontier since this will alter the historical distribution of returns. As a result, the efficient frontier cannot be reached for risk averse investors. And this efficient frontier should then result from a loser set of constraints, especially $w_i \geq 0$ and $w_1 + w_2 + w_3 + \cdots w_n = 1$, that should become (for example) $|w_1 + w_2 + w_3 + \cdots w_n| \leq Leverage_{ratio}$ and , $-w_{min} \leq w_i \leq w_{Max}$.

The nice feature here I that the constraints remain convex alongside the function that is to be optimized as long as the covariance matrix has all strictly positive eigen values. This last condition can be achieved in general through the method of perturbations.

Now if we assume a portfolio of "n" independent assets having each a mean return m_k and a variance v_k then the optimal weighting would secure a maximum ration $\frac{return}{standard_{deviation}}$

In other words we can look for the max of $\frac{\sum_{k=1}^{k=n} w_k m_k}{\sqrt{\sum_{k=1}^{k=n} w_k^2 v_k}} = R(w_1, w_2, .. w_n)$. We assume implicitly that the "risk" is best featured by the aggregate standard deviation of the return of the portfolio. To be sure, the "return" is the variation in price of the asset that has some "mean" value m_k over some fiwed period of time. It is not certain as such that we would realize this mean return unless we hold the asset up until some maturity time. The best example is for bonds. Another telling example is when one holds stocks for a long period of time: then the historical mean return tends to be realized. The "variance" that is used here reflect the statistical variance of the price. We do not need to assume that the prices follow a normal distribution. However most of the time we would rather consider the Ln(P) instead and consider the relative return on capital as a normally distributed variable. This gives more tractability to the model on the follow but imposes a different projection next for the P&L.

The optimum weighting secures that all the partial derivatives of this ratio are nil, given the constraint that $\sum_{k=1}^{k=n} w_k = 1$, hence we would only run on "n-1" independent coordinates ie "weights". Taking then $w_1 = 1 - \sum_{k=1}^{k=n} w_k$, we therefore have:

$$R(w_1, w_2, w_3 \ldots w_n) = \frac{\sum_{k=1}^{k=n} w_k m_k}{\sqrt{\sum_{k=1}^{k=n} w_k^2 v_k}}$$

Observe then that the weights appear as a square for the variance. It matters a lot for what follows. Indeed if the variance of a random variable X is "V", then the variance of the random variable "2X" will be "4V".

At the optimal weighting point using the 'n' dependent weights since the first is fully determined by them next we get for k=1,2,3….n

$$\frac{\partial R}{\partial w_k} = 0 = \frac{m_k \sqrt{\sum_{k=1}^{k=n} w_k^2 v_k} - \frac{1}{2}(2 v_k w_k) \overbrace{\frac{\sum_{k=1}^{k=n} w_k m_k}{\sqrt{\sum_{k=1}^{k=n} w_k^2 v_k}}}^{R_{opt}}}{\left(\sqrt{\sum_{k=1}^{k=n} w_k^2 v_k}\right)^2} = \frac{m_k - (v_k w_k) A}{\sqrt{\sum_{k=1}^{k=n} w_k^2 v_k}}$$

We infer then that $m_k = v_k w_k A \; with \; A = \frac{\sum_{k=1}^{k=n} w_k m_k}{\left(\sqrt{\sum_{k=1}^{k=n} w_k^2 v_k}\right)^2}$ and therefore $w_k = \frac{m_k}{v_k A}$

Do we have $\sum_{k=1}^{k=n} w_k = 1$?

Then $\sum_{k=1}^{k=n} w_k = 1 = \sum_{k=1}^{k=n} \frac{m_k}{v_k A}$ $and\ therefore\ A = \sum_{k=1}^{k=n} \frac{m_k}{v_k}$

As a result: $w_k = \frac{\frac{m_k}{v_k}}{\sum_{k=1}^{k=n} \frac{m_k}{v_k}}$

Then we could infer the optimal return based on these "n" independent assets that offer varying ratios $\frac{m_k}{v_k}$. Here the optimal ratio associated return and variances would respectively be:

$$m = \sum_{k=1}^{k=n} w_k m_k = \frac{\left(\sum_{k=1}^{k=n} \frac{m_k^2}{v_k}\right)}{\sum_{k=1}^{k=n} \frac{m_k}{v_k}} \ and\ v = \sum_{k=1}^{k=n} w_k^2 v_k = \frac{\left(\sum_{k=1}^{k=n} \frac{m_k^2}{v_k}\right)}{\left(\sum_{k=1}^{k=n} \frac{m_k}{v_k}\right)^2}$$

Thus the optimal ratio is going to be : $\frac{m}{\sqrt{v}} = \sqrt{\sum_{k=1}^{k=n} \frac{m_k^2}{v_k}}$. We can see that this solution takes the best of every component and beats every ratio $\frac{m_k}{\sqrt{v_k}}$ but it sets the level of "risk".

This calculation would only yield one "solution" for a portfolio of "n" assets where we would know in advance the expected return and associated variance.

What if, having a portfolio that carries 'n' different returns and variances, we wanted to know the optimal weights for a given target return? What if, instead, we wanted to know the best return for a fixed level of risk?

We need to identify the free variables of our problem then: the sum of the weights must be "1" so that we can compare the different possible configurations of the portfolio and the weighted return is fixed. Therefore $\sum_{k=1}^{k=n} w_k = 1\ and\ \sum_{k=1}^{k=n} w_k m_k = m$. Note that this works on the premise that indeed we can "know" for sure the ultimate return. The only case when this occurs is when we own "safe" bonds like government bonds AND we buy and hold them till they mature in full. Otherwise this premise is void of sense. But it is the most intuitive one to starts with.

We will set the assets by growing order of return and observe that only "n-2" weights are really free to fluctuate between 0% and 100%.

So the assets "1" and "2" will disappear in the optimization process, their respective weights being fully inferred from the other "n-2" assets.

We therefore state that : $w_1 + w_2 = 1 - \sum_{k=3}^{k=n} w_k\ and\ m_1 w_1 + m_2 w_2 = m - \sum_{k=3}^{k=n} w_k m_k$

We will note $\sum_{k=3}^{k=n} w_k = W,\ \sum_{k=3}^{k=n} w_k m_k = M\ and\ \sum_{k=3}^{k=n} w_k^2 v_k = V$ to simplify the resolution next.

Let's first "solve" $w_1\ and\ w_2$ relative to m,1,W and M:

$$\begin{cases} w_1 + w_2 = 1 - W \geq 0 \\ m_1 w_1 + m_2 w_2 = m - M \geq 0 \\ R = \dfrac{\sum_{k=1}^{k=n} w_k m_k}{\sqrt{\sum_{k=1}^{k=n} w_k^2 v_k}} \end{cases}$$

$$\begin{cases} -m_1w_1 - m_1w_2 = -m_1(1-W) \\ m_1w_1 + m_2w_2 = m - M \\ R = \dfrac{m_1w_1 + m_2w_2 + M}{\sqrt{w_1^2v_1 + w_2^2v_2 + V}} \end{cases} \quad ie \begin{cases} w_1 = (1-W) - w_2 \\ (m_2 - m_1)w_2 = m - M - m_1(1-W) \\ R = \dfrac{m_1w_1 + m_2w_2 + M}{\sqrt{w_1^2v_1 + w_2^2v_2 + V}} \end{cases}$$

$$\begin{cases} w_1 = (1-W) - \dfrac{m - M - m_1(1-W)}{(m_2 - m_1)} \\ w_2 = \dfrac{m - M - m_1(1-W)}{(m_2 - m_1)} \\ R = \dfrac{m_1w_1 + m_2w_2 + M}{\sqrt{w_1^2v_1 + w_2^2v_2 + V}} \end{cases} \quad ie \begin{cases} w_1 = \dfrac{m_2(1-W) - (m-M)}{(m_2 - m_1)} \geq 0 \\ w_2 = \dfrac{-m_1(1-W) + (m-M)}{(m_2 - m_1)} \geq 0 \\ R = \dfrac{m_1w_1 + m_2w_2 + M}{\sqrt{w_1^2v_1 + w_2^2v_2 + V}} \geq 0 \end{cases}$$

We find ultimately that “R” has a rather complex expression but one that is symmetrical for all the indices “k”:

$$R = \frac{m_1w_1 + m_2w_2 + M}{\sqrt{w_1^2v_1 + w_2^2v_2 + V}} = \frac{m_1\frac{m_2(1-W)-(m-M)}{(m_2-m_1)} + m_2\frac{-m_1(1-W)+(m-M)}{(m_2-m_1)} + M}{\sqrt{\left(\frac{m_2(1-W)-(m-M)}{(m_2-m_1)}\right)^2 v_1 + \left(\frac{-m_1(1-W)+(m-M)}{(m_2-m_1)}\right)^2 v_2 + V}}$$

We can simplify “R” first:

$$R = \frac{m_1w_1 + m_2w_2 + M}{\sqrt{w_1^2v_1 + w_2^2v_2 + V}} = \frac{m}{\sqrt{\left(\frac{m_2(1-W)-(m-M)}{(m_2-m_1)}\right)^2 v_1 + \left(\frac{-m_1(1-W)+(m-M)}{(m_2-m_1)}\right)^2 v_2 + V}}$$

We observe that the optimal point for “R” is exactly where “the total variance is nil”. So let’s simply look for this minimal total variance given that “m” is fixed anyway :

Let’s write $V_{ar} = \left(\frac{m_2(1-W)-(m-M)}{(m_2-m_1)}\right)^2 v_1 + \left(\frac{-m_1(1-W)+(m-M)}{(m_2-m_1)}\right)^2 v_2 + V$

The partial derivatives are therefore : knowing that $\frac{\partial W}{\partial w_k} = 1$, $\frac{\partial M}{\partial w_k} = m_k \; and \; \frac{\partial V}{\partial w_k} = 2w_kv_k$

Let’s write $V_{ar} = \left(\frac{m_2(1-W)-(m-M)}{(m_2-m_1)}\right)^2 v_1 + \left(\frac{-m_1(1-W)+(m-M)}{(m_2-m_1)}\right)^2 v_2 + V$

$$\frac{\partial V_{ar}}{\partial w_k} = 2v_1\left(\frac{m_2(-1) - (-m_k)}{(m_2 - m_1)}\right)\left(\frac{m_2(1-W) - (m-M)}{(m_2 - m_1)}\right) + 2v_2\left(\frac{-m_1(-1) + (-m_k)}{(m_2 - m_1)}\right)\left(\frac{-m_1(1-W) + (m-M)}{(m_2 - m_1)}\right) + 2w_kv_k$$

$$\frac{\partial V_{ar}}{\partial w_k} = 2v_1\left(\frac{m_k - m_2}{(m_2 - m_1)}\right)\left(\frac{m_2(1-W) - (m-M)}{(m_2 - m_1)}\right) + 2v_2\left(\frac{m_1 - m_k}{(m_2 - m_1)}\right)\left(\frac{-m_1(1-W) + (m-M)}{(m_2 - m_1)}\right) + 2w_kv_k$$

Which becomes :

$$\frac{\partial V_{ar}}{\partial w_k}$$
$$= 2\frac{\Big((1-W)\big(m_2v_1(m_k-m_2)-m_1v_2(m_1-m_k)\big)\Big)+(m-M)(v_2(m_1-m_k)-v_1(m_k-m_2))}{(m_2-m_1)^2}$$
$$+ 2w_kv_k$$

$$\frac{\partial V_{ar}}{\partial w_k} = 2\frac{\Big((1-W)\big(m_2v_1(-m_2)-m_1v_2(m_1)\big)\Big)+(m-M)(v_2(m_1)-v_1(-m_2))}{(m_2-m_1)^2}$$
$$+\frac{\Big((1-W)\big(m_2v_1(m_k)-m_1v_2(-m_k)\big)\Big)+(m-M)(v_2(-m_k)-v_1(m_k))}{(m_2-m_1)^2}$$
$$+ 2w_kv_k$$

$$\frac{\partial V_{ar}}{\partial w_k} = -2\underbrace{\frac{\left(\overbrace{(1-W)}^{C}(m_2^2v_1+m_1^2v_2)\right)-\overbrace{(m-M)}^{D}(v_2m_1+v_1m_2)}{(m_2-m_1)^2}}_{A}$$
$$+ m_k\underbrace{\frac{\left(\overbrace{(1-W)}^{C}(m_2v_1+m_1v_2)\right)-\overbrace{(m-M)}^{D}(v_2+v_1)}{(m_2-m_1)^2}}_{B}+2w_kv_k$$

We introduce values that are independent of "k" in order to "solve".

Since the optimum nullifies the partial derivatives for the free variables we infer that :

$$w_k = \frac{A-Bm_k}{v_k}$$

We will infer "C", "D" that we re-inject into A and B and finally find the values of A and B which will ultimately deliver the values of w_k and finally the values for w_1 and w_2

So: $C = 1-W = 1-\sum_{k=3}^{k=n} w_k = 1-\sum_{k=3}^{k=n}\frac{A-Bm_k}{v_k}$ and $D = m-M = m-\sum_{k=3}^{k=n} m_k\frac{A-Bm_k}{v_k}$

Therefore $\begin{cases} C = 1-A\sum_{k=3}^{k=n}\frac{1}{v_k}+B\sum_{k=3}^{k=n}\frac{m_k}{v_k} \\ D = m-A\sum_{k=3}^{k=n}+\frac{m_k}{v_k}B\sum_{k=3}^{k=n}\frac{m_k^2}{v_k} \\ A = \frac{\left(C(m_2^2v_1+m_1^2v_2)\right)-D(v_2m_1+v_1m_2)}{(m_2-m_1)^2}: \\ B = \frac{(C(m_2v_1+m_1v_2))-D(v_2+v_1)}{(m_2-m_1)^2} \end{cases}$

We should introduce then "known" values with $\begin{cases} S_0 = \sum_{k=3}^{k=n}\frac{1}{v_k} \\ S_1 = \sum_{k=3}^{k=n}\frac{m_k}{v_k} \\ S_2 = \sum_{k=3}^{k=n}\frac{m_k^2}{v_k} \end{cases}$ and $\begin{cases} r_2 = \frac{(m_2^2v_1+m_1^2v_2)}{(m_2-m_1)^2} \\ r_1 = \frac{(m_1v_2+m_2v_1)}{(m_2-m_1)^2} \\ r_0 = \frac{(v_2+v_1)}{(m_2-m_1)^2} \end{cases}$

The indices point to the power law of the means….

The system to solve for "A,B,C,D" then becomes:

$$\begin{cases} C = 1 - AS_0 + BS_1 \\ D = m - AS_1BS_2 \\ A = Cr_2 - Dr_1 \\ B = Cr_1 - Dr_0 \end{cases} \text{ therefore: } \begin{cases} C = 1 - AS_0 + BS_1 \\ D = m - AS_1 + BS_2 \\ A = r_2(1 - AS_0 + BS_1) - r_1(m - AS_1 + BS_2) \\ B = r_1(1 - AS_0 + BS_1) - r_0(m - AS_1 + BS_2) \end{cases}$$

$$\begin{cases} C = 1 - AS_0 + BS_1 \\ D = m - AS_1 + BS_2 \\ A = r_2 - r_1 m - A(r_2S_0 - r_1S_1) + B(r_2S_1 - r_1S_2) \\ B = r_1 - r_0 m - A(r_1S_0 - r_0S_1) + B(r_1S_1 - r_0S_2) \end{cases}$$

We finally solve the respective values for A and B:

$$\begin{cases} A = \dfrac{r_2 - r_1 m + B(r_2S_1 - r_1S_2)}{1 + (r_2S_0 - r_1S_1)} \\ B = r_1 - r_0 m - \dfrac{r_2 - r_1 m + B(r_2S_1 - r_1S_2)}{1 + (r_2S_0 - r_1S_1)}(r_1S_0 - r_0S_1) + B(r_1S_1 - r_0S_2) \end{cases}$$

$$\begin{cases} A = \dfrac{r_2 - r_1 m + B(r_2S_1 - r_1S_2)}{1 + (r_2S_0 - r_1S_1)} \\ B\left(1 - (r_1S_1 - r_0S_2) + \dfrac{(r_2S_1 - r_1S_2)}{1 + (r_2S_0 - r_1S_1)}\right) = r_1 - r_0 m - \dfrac{(r_2 - r_1 m)(r_1S_0 - r_0S_1)}{1 + (r_2S_0 - r_1S_1)} \end{cases}$$

$$\begin{cases} A = \dfrac{r_2 - r_1 m + B(r_2S_1 - r_1S_2)}{1 + (r_2S_0 - r_1S_1)} \\ B = \dfrac{r_1 - r_0 m - \frac{(r_2 - r_1 m)(r_1S_0 - r_0S_1)}{1 + (r_2S_0 - r_1S_1)}}{\left(1 - (r_1S_1 - r_0S_2) + \frac{(r_2S_1 - r_1S_2)}{1 + (r_2S_0 - r_1S_1)}\right)} \end{cases}$$

We can see that we can mechanically infer the weights because we can compute the values for A and B , the C and D, and ultimately the values for the weights for k≤ 3 with $w_k = \frac{A - Bm_k}{v_k}$ and infer finally the other weights with :

$$\begin{cases} w_1 = \dfrac{m_2(1 - W) - (m - M)}{(m_2 - m_1)} \geq 0 \\ w_2 = \dfrac{-m_1(1 - W) + (m - M)}{(m_2 - m_1)} \\ R = \dfrac{m_1 w_1 + m_2 w_2 + M}{\sqrt{w_1^2 v_1 + w_2^2 v_2 + V}} \geq 0 \end{cases} = \begin{cases} w_1 = \dfrac{m_2 C - D}{(m_2 - m_1)} \geq 0 \\ w_2 = \dfrac{-m_1 C + D}{(m_2 - m_1)} \\ R = \dfrac{m_1 w_1 + m_2 w_2 + M}{\sqrt{w_1^2 v_1 + w_2^2 v_2 + V}} \geq 0 \end{cases}$$

Now we could adopt an alternative approach saying: we fix the variance and look for the optimal expected return for this given aggregate variance.

The calculation is following the same principles. It has the advantage here that we do not need to explicitly assume that we “know” the ultimate return. As mentioned, this part can be quite problematic because the former premise that we would know the final return imposes that we only invest in fixed income products, that we ignore the default risk and that we hold the positions until maturity. Instead, assuming that we fix the variance and look for some optimal return then we introduce a more opportunistic part premising then that we only need to know the expectation of the return and next act accordingly all the way. Yet this introduces a review of the notion of

"expected return" and above all the "expectation of the maximum return". See the appendix for details on this approach that yet is more realistic.

But here we assumed that the variables were independent and normally distributed. Also we quite importantly we need to review what we should do when, and this is the case in general, the returns are more or less correlated. We can always reduce the covariance matrix with a basis of eigen vectors and therefore solve the problem in the eigen basis where, statistically, the variables now are uncorrelated which brings us back to the precedent development.

We will stick to the assumption that we "know" the expected return. Note that if we work with "expected maximum returns" all this optimization works equally well.

There is another way, more geometrical, to compute the frontier by defining first the eigen values and the eigen vectors. Indeed C, the covariance matrix being real and symmetrical, we can find a basis of eigen vectors ie we can transpose the problem to a case where we would only deal with n "independent" "eigen assets" based upon the coordinates of the eigen vectors that would diagonalize the empirical covariance matrix. Note that this approach can be generalized to higher order derivatives : we can either generate a gradient for odd degrees or we can combine pairs of degrees. We would not obtain always obtain perfectly diagonal matrices but, assuming that we always deal with smooth functions we would always be able to invert the order of derivation freely.

Sticking with the core model here in 2 dimensions we therefore could rephrase the problem in the eigen basis as: (note that we have a different set of factors for the last constraint that will depend upon the coordinates of the eigen vectors when each original weight would be expressed, as a vector, through a linear combination in the eigen basis).

Introducing the aggregate factors $(a_1, a_2 \dots a_n)$ that move us back from the eigen basis into the canonical basis, we look at the following optimization problem (our eigen weights may not match a real life situation where the weights in the canonical basis sum up to 1: $v_w =$

$$\sum_{k=1}^{k=n} v_{pk}\, w_{eige\ \ _k}^2 \; under\ the\ constraints: \begin{cases} {w_{eigen}}^T r_{eigen} = r_{fixed} \\ w_{eigen_i} \geq 0 \\ a_1 w_{eige\ \ _1} + a_2 w_{eigen_2} + a_3 w_{eige\ \ _3} + \cdots a_n w_{eign\ \ _n} = 1 \end{cases}$$

By changing the basis of representation one more time, moving towards $y_k = \sqrt{v_{pk}} w_{eigen_k}$ we land onto a very geometrical solution:

Minimize

$$R_y^2 = \sum_{k=1}^{k=n} y_k^2 \; under\ the\ constraints: \begin{cases} \frac{r_1}{r_{fixed}} \frac{y_1}{\sqrt{v_{p1}}} + \frac{r_2}{r_{fixed}} \frac{y_2}{\sqrt{v_{p2}}} + \frac{r_3}{r_{fixed}} \frac{y_3}{\sqrt{v_{p3}}} + \cdots \frac{r_n}{r_{fixed}} \frac{y_n}{\sqrt{v_{pn}}} = 1 \\ y_k \geq 0 \\ a_1 \frac{y_1}{\sqrt{v_{p1}}} + a_2 \frac{y_2}{\sqrt{v_{p2}}} + a_3 \frac{y_3}{\sqrt{v_{p3}}} + \cdots a_n \frac{y_n}{\sqrt{v_{pn}}} = 1 \end{cases}$$

So the optimal problem becomes a rather simple geometrical question in n dimensions : find the smallest sphere that intersects 2 hyperplanes as defined in the constraints, in the « positive quadrant ». We studied in detail how one can solve it explicitly but at the expense of tedious calculations. Here, once diagonalized, we can adopt a numerical resolution based on this geometrical view.

Then we could eliminate one coordinate so that we remain with only one condition on the "n-1" remaining coordinates and therefore only one hyperplane in some "n-1" space. The "squared radius"

function would become some elliptic function then. But we can always change one last time the coordinates so that to retrieve once more a squared radius equation and infer the radius of the tangent circle along with the coordinates of the tangent point

Let's consider a naïve portfolio that truly is the actual sum of the "n" native assets and let's see how the "dependence function" that should be independent from the sizes involved pops up from, first, the joint distribution, and second, from our usual target that is an expectation computed in the middle of a distribution of outcomes.

So the "closed formula" can be obtained once one diagonalizes the covariance matrix and expresses the native portfolio as a portfolio of "eigen" assets instead where one has to compute the associated expected return and expected variance, assuming on the follow that we have a normal distribution.

We can also adopt a more numerical resolution using a geometrical interpretation. But what if an investor is more risk averse and actually cannot really bear any level of risk for the sake of maximizing its risk reward? What if an investor is constrained by the range of risks it can take? How should we proceed? Is the variance actually the only "risk" for an investor. The answer is "no": an investor mostly fear the "tail risk", the "draw-down", where a lasting large loss has to be shouldered if the investor wants ultimately to be paid for the risk that was undertaken at the start of all this investment strategy. So we need to adopt a more comprehensive framework that would include higher statistical moments.

The short answer is we maximize a weighted return under a Lagrange constraint. Let's take a very theoretical approach as to what this entails.

So we would write the expected return of such portfolio as :

$$P(X_1 + X_2 \dots + X_n) = (x_1 + x_2 + x_3 \dots)$$

And we next should consider the many consecutive related moments since the n random variables can take many values as per their individual distributions and as per the joint distribution that binds the "n" independent variables together.

$$M_k(P) = \iiint (x_1 + x_2 + x_3 \dots)^k \prod_{k=1}^{k=n} p_k(X_k = x_k) dx_k$$

We know that the variables $X_1, X_2, X_3 \dots X_n$ are random and follow their own distribution p_k.

Remember: the formula above implicitly assumes that the variables are independent. If the variables are not independent, the theorem of Sklar introduces the idea that, should the n random variables be of same magnitudes on the first 2 moments, there is a "copula C" that allows us to define the joint distribution.

So, let's first write "the obvious" ie that "yes" we have a joint distribution even when the n random variables are not independent. Then the formula above should be written as follows:

$$M_{P_k}(X_1 + X_2 \dots + X_n) = \iiint (x_1 + x_2 + x_3 \dots)^k P_{joint}(X_1 = x_1, X_2 = x_2 \dots) \prod_{k=1}^{k=n} dx_k \dots$$

The theorem of Sklär allows us to write, quite importantly the following equivalence:

$$M_{P_k}(X_1 + X_2 \dots + X_n) = \iiint (x_1 + x_2 + x_3 \dots)^k \, C(\{p_k(X_k = x_k)\ k = 1,2, \dots n\}) \prod_{k=1}^{k=n} dx_k \dots$$

The very existence of this "Copula", at least from a pure theoretical standpoint that mostly is valid when we adhere to the existence of a unique class of joint distributions, has a very important consequence: since we would say that whatever the multiplier a_k that we may apply to the variable X_k we should certainly assume that $P(a_k X_k = a_k x_k) = P(X_k = a_k)$. Therefore the same copula would apply to any "weighting" that we may use on the native assets. Note however that the equality is not always valid if the "size matters" on some assets.

We should think of these variables less as "price" and more in terms of "returns" in relative values. Thus the expectation would be in the order of magnitude of a few % points in most cases but it could reach extreme values in exceptional circumstances. In this fashion we likely would get close enough to a normal distribution.

Intuitively, the primary target is the highest return, which in our framework points towards the maximum moment of first order. However, investors would have a "risk aversion" pattern that is expressed in the simplest form through the "risk" that they would undertake in "volatility" terms. This would generally correspond to the "variance", ie the squared distance relative to the expectation.

The variance, with our notations, would be expressed as follows:

$$V = \iiint \left(x_1 + x_2 + x_3 \dots - M_{P_1}(X_1 + X_2 \dots + X_n)\right)^2 C(\{p_k(X_k = x_k)\ k = 1,2, \dots n\}) \prod_{k=1}^{k=n} dx_k$$

Since $M_{P_k}(X_1 + X_2 \dots + X_n) = \iiint (x_1 + x_2 + x_3 \dots)^k \, P_{joint}(X_1 = x_1, X_2 = x_2 \dots) \prod_{i=1}^{i=n} dx_k$

When we would develop the "square" that is present under the integral signs, we would obtain a classical result:

$$\begin{aligned} V = \iiint (x_1 + x_2 + x_3 \dots)^2 \, C(\{p_k(X_k = x_k)\ k = 1,2, \dots n\}) \prod_{k=1}^{k=n} dx_k \\ - 2M_{P_1}(X_1 + X_2 \dots + X_n) \iiint (x_1 + x_2 + x_3 \dots)^1 \, C(\{p_k(X_k = x_k)\ k \\ = 1,2, \dots n\}) \prod_{k=1}^{k=n} dx_k + M_{P_1}(X_1 + X_2 \dots + X_n)^2 \end{aligned}$$

And we would infer that : $V = M_2(X_1 + X_2 \dots + X_n)^1 - M_{P_1}(X_1 + X_2 \dots + X_n)^2$

The investors being risk-averse would ponder the benefit against the "cost", ie the expected return against the risk that, because of the variance, they may go through a rough patch and then be forced to close prematurely their investment at a loss.

The most intuitive angle to it consists in trying to optimize the expected return with a penalty that would come from the associated value of the variance:

$$F(portfolio) = Max[M_{P_1}^2 - \beta M_{P_2}]$$

In our earlier calculations we made quite strong assumptions. First we assumed that an investor "knows" what the ultimate return will be. Well that is true for "buy and hold" investors who buy

bonds or else buy shares but over the very long term without changing much their strategy. This is actually the exception rather than the rule. Most investors have or fell the duty to actually "optimize" their money management, which implies that they monitor the events, the news, the price changes, the resulting P&L and returns and…make decisions to change the balance of risks in their portfolio. Doing so, they usually "cut losses" or "add risks" or else "take profits". In general they face constraints as to the envelope of risks that they are allowed to take by mandate. Therefore it would be more intuitive to actually look to maximize a return over a certain predefined period with the constraint that the "risk" is fixed. But is the risk "fixed" in essence? The answer is "no": the risk is volatile in essence. So, the framework is more the one where the risk is rather random and the return accordingly is more or less volatile. The whole now is to assume that "no" investors do not in general target the Sharpe ratio other than on rather short periods of time and under the constraint that they have a risk-aversion pattern. Thus investors look for an efficient frontier indeed but more in the way that we will picture in what follows.

We look to optimize the following function:

$$F(w_1, w_2, .. w_n) = Return(w_1, w_2, ... w_n) - \beta Risk(w_1, w_2, .. w_n)$$

The "risk" is typically measured with the standard deviation but it might include higher moments that feature the "tail risk" or else higher statistical moments with an additional weight.

We immediately note that the optimal configuration may depend upon the choice of our penalty β: the higher β is the lower the expected return. But what is the reasonable β given a certain level of risk? How is this β penalty changing when the risk aversion for the investor becomes lower than it was? We sense that for a given set of "n" assets, given their distributions, given their joint distribution, the penalty β varies. This means that this also must be re-assessed at regular intervals.

That would make sense since the best return that one can expect should grow along with the appetite for risk, the risk is varying and the investor's appetite for risk also varies independently so. This is mandated by the fact that, on balance, the higher the risk we take, the higher the probability that we undergo an underperformance in the end. In other words there should be no free-lunch: a risk averse investor should not be able to expect a higher return and in exchange for that he will get a more predictable but low ultimate outcome. What should we do, if we know the level of risk that we can accept, in order to spot the "best possible return", ie how do we choose the "best possible penalty through the choice of β?

This approach, if we consider that the choice of portfolio is made by picking the best weights among "n" assets, is close to the one that is implemented through the computation of the efficient frontier. In the latter case, we would spot the optimal set of weights for any given target return by searching for the configuration of the portfolio that would induce the smallest variance. So, once we would know the appetite for risk of an investor, we would look at this efficient frontier, and spot the maximum expected return we could get. And on that specific point of the efficient frontier, we would know the optimal weights that should deliver both the target return that we would inferred from the stated "tolerated variance" number.

Looking back at the last formula above, we could see the connection: if we look at the derivative of the formula relative to M_{P_2}, a risk neutral move would induce that $\partial_{M_{P_2}}[F(Portfolio)] = 0$. This would amount to say that $\partial_{M_{P_2}}\left[M_{P_1}^2\right] = \beta$ which is more or less the slope of the efficient frontier for a certain level of risk M_{P_2} . So, if we start wondering "what should be the optimal penalty that we should impose on risk given a give level if we want to achieve the best possible rate of return?", the answer would be: "take the slope of the efficient frontier".

We see here what the concept of “risk-neutrality” conveys: people may have varying risk appetite but a “risk-neutral” approach to risk itself would consist in moving along the efficient frontier. But, given that there are other determinants of the risk aversion profile of investors that may include other moments, and even more specific “path dependent” limitations, how should we model and compute (easily?!) the equivalent “efficient frontier”?

One quick review of the Sharpe ratio approach brings in a little more context. In a Sharpe ratio approach we actually consider the following optimization on the function:

$$S_{har} \quad (portfolio) = Max\left[\frac{M_{P_1}^2}{M_{P_2}} - \beta\right] = Max\left[\frac{M_{P_1}^2}{M_{P_2}}\right]$$

If we contrast the 2 formula, we observe a simple relation between the 2:

$$F(portfolio) = Max\left[M_{P_1}^2 - \beta M_{P_2}\right] = M_{P_2} S_{harp} \ (portfolio)$$

Clearly, the spontaneous concavity of an efficient frontier in general will induce a choice to target the lowest possible risks in variance terms and create next a maximum financial leverage effect over this tiny but ultra safe return. The ongoing concavity of the efficient frontier signals indeed that the slope of tangent to the efficient frontier can only decline. In more intuitive terms, this only says that any fixed increment of risk will induce a lower and lower gain in expected returns. As a result the best “slope” is found at “risk=0+”. Then a player might generate a multiplier effect by borrowing a lot of money in order to create a high multiple of the native return undertaking mostly counterparty risks.

This is what banks do and this is what tends to generate the so called “systemic risk”. The original reason why banks do that comes from the search of “safety” for the savings and deposits that induce very strict regulations. However, this need for “safety” does not fit with the conflicting need for banks to deliver the highest possible return on capital to their shareholders. They therefore mitigate the 2 contradicting objectives through this strategy that consists in targeting “AAA like” ie low returns but with the bare minimum risk. Since they need to deliver high returns banks borrow a maximum amount of money and build financial leverage around this arbitrage. At the other end of the spectrum we find “distress funds”, but also “pension funds” or “insurers”, or “corporate” that benefit in small parts from regulation exemptions and target the maximum risk and minimum financial leverage.

As such this leveraged approach could and should be modeled under a similar framework considering that “one asset” is the safe portfolio and another asset is the “ability to borrow collateralized money”. Of the 2 “assets” one is triggering financial crisis at times and thus one may as well opt to look for “diversified” sources of portfolios and “diversified sources of external” funding in order to alleviate the risk here. But so far our financial markets are centralized around one unique US$ ring of funding across the planet. Crypto currencies, the Yuan are recent alternatives but they do not have by far the same size and depth for now.

What if we aim to include other moments in this approach in order to express, for example, some “shortage” function?

Investors might first target the highest Sharpe ratio (expectation/standard deviation) while they actually may not whither properly some draw-downs or else stick to a strategy that has “won already”. The second alternative, ie when the portfolio stands at some outstanding gain, seems “too good” to be true: it may have perverse consequences actually. Indeed, as much as we can easily conceive the scenario where an investor faces temporarily a loss that is “much higher than expected”, ie a multiple of the standard deviation that points to some “ab-normally distributed”

outcome: in that case it would prompt an unwanted wind down of the investment and a fixed loss. We would consequently figure that the “super nice and unexpected profits” would be the reciprocal situation. Then an investor might “opt out” to secure these windfall gains. However, this situation induces choices on the follow that may have perverse consequences. The main issue then being “what to do next?”. Since the original strategy is exited, the investor would have to choose “something else”. Here this would be a deviation from the initial strategy that would be mandated by some opportunistic decision process. And, although the timing of the exit would prove to be right, whatever decision that would be made next would be an additional risk.

One can start e by including the “Skew” (moment of 3rd degree) and the “Kurtosis” (moment of 4th degree) of the distribution of returns in order to factor in these aspects. The concept underneath is the ” shortage function” that embodies this view that “mean” and “variance” do not fit with the actual context of an investor, irrespective of its “risk aversion” pattern. Indeed investors have varying patterns, aiming at different returns and therefore readying themselves to different levels of risk. However, their issue is, as much, about the sequencing of their investment process. Why is that? Because in reality the path that will be followed will matter, irrespective of whether these investors are happy to “take profits” or “stop out” or else “strategize” at regular intervals. They merely have to do it anyway, albeit in the form of some algorithmic investment process.

So the question now bears on the ability to define an equivalent “efficient space” that would embed more risk factors that the simple “variance” versus the “expected return”.

At first we need to clarify the relation between the copula and the traditional approach when all is based on the empirical covariance. Typically, once we have reduced the estimation of the moment of second order to the covariance matrix we often make this observation that this covariance itself can be split into a correlation matrix that is independent of the “sizes involved” and a vector of variances that will carry the “sizes involved”.

Can we generalize this approach that is typical once we assume that we deal with normal distribution and therefore a normal correlated joint distribution of “n” variables?

How is this possible with a copula? Is this a wise move?

First let’s clarify the core assumptions here…

The correlation matrix empirically is established as follows

$$C_{orr}(X_1, X_2, \dots X_n) = \left[C_{or\ ij} = \frac{\frac{1}{N-1}\sum_{k=1}^{N}\left(X_{i_k} - X_{i_{mean}}\right)\left(x_{j_k} - X_{j_{mean}}\right)}{\sqrt{\frac{1}{N-1}\sum_{k=1}^{k=N}\left(X_{i_k} - Xi_{mean}\right)^2}\sqrt{\frac{1}{N-1}\sum_{k=1}^{k=N}\left(X_{j_k} - X_{j_{mean}}\right)^2}} \right]$$

We can then define a vector-column of standard deviations:

$$S_{td} = \left[S_{td_i} = \sqrt{\frac{1}{N-1}\sum_{k=1}^{k=N}\left(X_{i_k} - Xi_{mean}\right)^2} \right]$$

And of course we would retrieve the covariance matrix through a simple formula:

$$C_{ov}(X_1, X_2, \dots X_n) = \begin{bmatrix} S_{td_1} \\ S_{td_2} \\ S_{td_3} \\ \dots \\ S_{td_n} \end{bmatrix} C_{orr}(X_1, X_2, \dots X_n) \begin{bmatrix} S_{td_1} & S_{td_2} & S_{td_3} & \dots . & S_{td_n} \end{bmatrix}$$

Observe that this is not an "outer product" but close to that: the column vector is n the right and the line vector is on the left which implies that there is no "summation", but just simple products. This is useful in programming terms. Thanks to the symmetrical nature of the correlation matrix we could express this in a basis of unit norm eigen vectors and there the correlation matrix would be a diagonal matrix. As such it is the reverse to what we usually use to compute a scalar from a bilinear form.

The expression has an elegant property: the "dependence" is expressed separately from the "size" of the components of the "sum of random variables" that is at play here.

Could we retrieve this elegant property from an expression like, knowing that the covariance matrix is the empirical estimate of the moments of order 2 in "n" dimension?

$$P_{joint}(X_1 = x_1, X_2 = x_2 \dots) = C(\{p_k(X_k = x_k)\ k = 1,2, \dots n\})$$

We need to formulate the standard deviation, the covariance matrix with the introduction of the standalone probabilities. We saw that the covariance element is, from its empirical definition towards its theoretical definition, defined as:

$$\frac{1}{N-1}\sum_{k=1}^{N}(X_{i_k} - X_{i_{mean}})(X_{j_k} - X_{j_{mean}}) \approx \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x_i - x_{i_{mean}})(x_j - x_{j_{mean}})C(p_{x_i}, p_{x_j})dx_i dx_j$$

In order to better see the connection here, let's imagine that with the "N" empirical data we actually could split the samples into some p² sub-intervals that would cover the whole range of possible values for X_i on the one hand and for X_j on the other hand. We could then "count" the number of elements of each variable respectively in each sub-interval. Then we could write the approximation:

$$\frac{1}{N-1}\sum_{k=1}^{N}(X_{i_k} - X_{i_{mean}})(X_{j_k} - X_{j_{mean}}) \approx \frac{1}{p^2}\sum_{k_1=1}^{k_1=p}\sum_{k_2=1}^{k_2=p}\left(x_{i_{k_1}} - x_{i_{mean}}\right)\left(x_{j_{k_2}} - x_{j_{mean}}\right)p_{x_{i_{k_1}}, x_{j_{k_2}}}$$

We now can see where we come from applying the theorem of Sklär that states that there is a joint distribution that we name "copula" under rather loose conditions:

$$p_{x_{i_{k_1}}, x_{j_{k_2}}} \approx C\left(p_{x_{i_{k_1}}}, p_{x_{j_{k_2}}}\right)dx_i dx_j$$

The probability, or should we rather say the "density of probability" $p_{x_{i_{k_1}}, x_{j_{k_2}}}$, is what the copula embodies which points to the fact the Copula remains a function of the "bucket" that we focus on here even though the can set a function that purely depends on the the value of this density. We see here that IF the 2 variables were locally independent in a bucket we then ALSO could write that

$$p_{x_{i_{k_1}}, x_{j_{k_2}}} = p_{x_{i_{k_1}}}\, p_{x_{j_{k_2}}} \approx C\left(p_{x_{i_{k_1}}}, p_{x_{j_{k_2}}}\right)dx_i dx_j$$

Then we would infer that C is the identity density function as a function of the individual probabilities. In the more general case, we should write:

$$p_{x_{i_{k_1}},x_{j_{k_2}}} = \rho_{x_{i_{k_1}},x_{j_{k_2}}} p_{x_{i_{k_1}}} p_{x_{j_{k_2}}} \approx C\left(p_{x_{i_{k_1}}}, p_{x_{j_{k_2}}}\right) dx_i dx_j$$

The factor, at first order, feels very much like a "correlation factor" but it represents the dependence relation in a different fashion: it quantifies the chance that only the individual probabilities sufficiently describe the conjunction of co-presence in each bucket. We can observe that from a numerical standpoint we would only care about "p" which would be anyway independent of the eventual "weight" of any asset. We formerly hinted at the "density of probability"…Well here the density of probability is $C\left(p_{x_{i_{k_1}}}, p_{x_{j_{k_2}}}\right) \approx \rho_{x_{i_{k_1}},x_{j_{k_2}}} \frac{p_{x_{i_{k_1}}}}{dx_{i_{k_1}}} \frac{p_{x_{j_{k_2}}}}{dx_{j_{k_2}}}$

Of course the real density of probability would be "lower" as we grow the "granularity" of our bucketing with which we would "cut" the empirical ranges into sub-intervals.

Let's observe, and this is where the "copula" concept is appealing: this model assumes that the other terms do not matter ie $C\left(p_{x_{i_{k_1}}}, p_{x_{j_{k_2}}}\right)$ is NOT conditioned by the values that are taken by the other p_{x_m} with "m" differing from "j and i". We can see here how the concept generalizes what we usually assume with the normal distribution and a "fixed" dependence matrix. This simply comes from the fact that a normal distribution is fully determined by its mean and variance (maximum entropy distribution through a range of $]-\infty,+\infty[$ $for\ a\ known\ mean\ and\ variance$).

One powerful property of the Copula concept is that, if we can keep the same value for "p" irrespective of the "scale" of the weight that would be attributed to each asset, we then deal with constant values relative to the dependence effects.

It feels like ,so far, that we "only" deal with 2 variables at a time. This is not quite true since "C" defines the interdependence of "n" variables across their respective ranges. And the dependence going way beyond just "pairs" of variables, we must consider that $\rho_{x_{i_{k_1}},x_{j_{k_2}}}$ is a function of all the "n" buckets through the other variables. In other words, the theorem of Sklär states that if the dependence factor for 2 given variables depends upon the value of the others, only the probabilities would matter for the other variables since, as stated, this Copula function does exist that is a function of these probabilities alone, not the value of the variables themselves. This covers any distribution that is sufficiently well behaved for sets of variables that are sufficiently homogeneous, a thing that we can easily assume here.

Now, the whole issue is whether the empirical evidence is large enough and, more importantly, whether we can infer a reliable dependence factor numerically speaking IF we are not dealing with proxy joint normal distributions in the background. This is where we face a problem: only the normal distribution provides a dependence factor that can be estimated reliably from empirical evidence and that matches reliable interpolations or extrapolations for prediction purpose. However, assuming for example that we have a sum of normal distributions that have different "spikes", we can work with that and refine the correlation estimates by making them "conditional on the other probabilities of presence".

So from the usual empirical proxy that we apply with the covariance matrix, we actually "miss" the fact that the correlation factor may itself be a function of the other variables and that this sole aspect renders the empirical traditional estimate of the "variance" a misleading one. This is a problem that surfaced with credit derivatives and mandated the use of Copula based models. Even so, the calibration did not work well for other more fundamental reasons like: there is no "risk neutral" rate

and there is no market equilibrium. Still this approach is useful for portfolio risk management (even if it is not sufficient for pricing purpose).

How can we secure that standpoint and, further, how can we introduce in practice a standard deviation value that would permit to separate the "size" through this standard deviation from the "correlation pattern" seen now as a multi-dimensional "dependence pattern"?

Back to the root equation for order 2, we wrote initially:

$$\frac{1}{N-1}\sum_{k=1}^{N}\left(X_{i_k}-X_{i_{mean}}\right)\left(X_{j_k}-X_{j_{mean}}\right)\approx\frac{1}{p^2}\sum_{k_1=1}^{k_1=p}\sum_{k_2=1}^{k_2=p}\left(x_{i_{k_1}}-x_{i_{mean}}\right)\left(x_{j_{k_2}}-x_{j_{mean}}\right)p_{x_{i_{k_1}},x_{j_{k_2}}}$$

We simply stated above that the common "correlation" is some "average" of the dependence and this average itself is averaging other "averages" over the other distributions of the other variables entering in the aggregate joint distribution. Here, thanks to the theorem of Sklär, under rather loos constraints, we have a "weight" invariant measure of the dependence as long as we can "break" the range of possible values into a fix number of sub-intervals "scale-wise". Thus if we were to change the weights we would adjust automatically in proportion the size of the sub-intervals which would guarantee that the probability $p_{x_{i_{k_1}},x_{j_{k_2}}}$, but also the standalone probabilities $p_{x_{1_{k_1}}}\ and\ p_{x_{2_{k_2}}}$, would remain unaltered, and so would the correlation local factor $\rho_{x_{i_{k_1}},x_{j_{k_2}}}$.

Above we observe , that with these conventions in place, the "weights" effect associated to some "unit standard deviation" could be equally introduced through the terms $\left(x_{i_{k_1}}-x_{i_{mean}}\right)p_{x_{1_{k_1}}}$ since at the end of the day we could estimate a "unit +1" standard deviation based upon

$$S_{td_{x_i}}=\sqrt{\frac{1}{p^2}\sum_{k_1=1}^{k_1=p}\sum_{k_2=1}^{k_2=p}\left(x_{i_{k_1}}-x_{i_{mean}}\right)\left(x_{1_{k_2}}-x_{i_{mean}}\right)p_{x_{1_{k_1}}}\,p_{x_{1_{k_1}}}}$$

But here we would carry a very gross estimate of both the local standard deviation and the local correlation pattern that exists between the 2 standalone probabilities in ever bucket.

Yet we can now bridge the gap between the "local" view that is brought up by the concept of Copula and the standard empirical estimate of the "variance":

$$\frac{1}{p^2}\sum_{k_1=1}^{k_1=p}\sum_{k_2=1}^{k_2=p}\rho_{x_{i_{k_1}},x_{j_{k_2}}}\left(x_{i_{k_1}}-x_{i_{mean}}\right)p_{x_{1_{k_1}}}\left(x_{j_{k_2}}-x_{j_{mean}}\right)\,p_{x_{2_{k_2}}}=\rho_{x_i,x_j}S_{td_{x_i}}S_{td_{x_j}}$$

The emphasis on the weights being "specified" or not specified will show its importance later on when we will aim at shaping the "risk profile" of the investor n the one hand, and search the "optimal weights" next on the other hand. So far we have established how we could define "local correlation factors and how they relate the otherwise well chartered "linear correlation" factor. We above how this factor is a mere expectation, in 2 dimensions, of local standard deviations themselves averaged.

Can we extend this approach to other moments of higher order, like the "skew" , the "kurtosis" and beyond to any order? If "yes" is there a link towards the "moment generating" function in n dimensions?

We have now to fill the gap between this correlation pattern, including a copula or being just an average correlation factor across the range of variations, and the optimization target. The latter would most spontaneously show up an optimal return, risk adjusted. And here we could invoke the model of Harry Markowitz with the "efficient" frontier: for a given level of risk there is a maximum return that corresponds to a certain set of weights. These weights are inferred from the "slope" of the tangent line to the efficient frontier at the absciss that is equal to the risk level. The risk in itself is estimated with the variance against the squared return provided the variables are "independent". Equivalently we could optimize the expected return versus the standard deviation. The slope would change but the principle remains the same. We also saw that in general the variables are NOT independent. Yet we can use a correlation matrix, diagonalize it and revert to a formulation of the problem that is using independent "eigen" assets. We just saw that most of the time we can refine this approach further provided we have enough empirical data, define a Copula model and run this optimization conditioned upon the current probabilities of the other variables. It is consistent. But so far we only dealt with the first 2 statistical moments.

6- Time series , diagonalization and LLMs

This part deals with the basic techniques that handle time series and linear regressions.

The first part sketches the core issues for financial modeling and introduces the concept of "spectral analysis" where one focuses on the frequencies rather than the trends.

The second part puts an emphasis on the diagonalisation of matrices

The Last part summarizes the principles that govern LLMs like ChatGPT

The first part: trends and frequencies…..

Warm-up: exercises

1- given m and n 2 integers with a ppcm a and a pgcd b: find a simple expression of "ab" with m and n
2- Let a polynomial of 2nd order P. We know one root x and the value y where its derivative is 0: What is the value of the second root as a function of x and y?
3- What is the simple expression for $1 + x + x^2 + x^3 + \cdots . when\ |x| < 1$?
4- What is the value of $\sum_{k=1}^{N} k$? What is the value of $\sum_{k-1}^{N} k^2$? What is the value of $\sum_{k=1}^{N} k^3$?
5- Taylor-Young formula in 1 dimension: derivation, approximation… intuition….: any time series can be approximated with a polynomial with an arbitrary accuracy does it make sense?

Course

Executive summary:

A linear regression renders a simplified but powerful local "solution" for interpolation and extrapolation. It is useful but very soon it becomes frustrating and inefficient. Once one gets a relatively nicely correlated cloud of points, one should expect the CLT (Central Limit Theorem) to work its magic: if you assume that the dispersion from the mean ultimately should follow a normal distribution, then betting on the mean reversion mechanism should always work at one point in time. Unfortunately this is true only "at the end of times" and, in the meantime, the mean and the