Overcoming Markowitz's instability with the help of the Hierarchical Risk Parity: theoretical evidence.

Alexandre Antonov, Alexander Lipton and Marcos López de Prado

ADIA

March 23, 2023

Introduction

- In this presentation we compare two methods of portfolio allocation: the classical Markowitz one [2] and the hierarchical risk parity (HRP) based on a clustered optimization [4]
- We drive theoretical values of a noise of the allocation weights coming from the covariance matrix numerical estimation.
- We demonstrate that the HRP is indeed less noisy (and thus more robust) w.r.t. the classical Markowitz
- We confirm the theory using Monte Carlo simulations
- We have derived the formulas for the min-variance optimization, Gaussians assets and low cross-correlations. However, our results can be generalized for other (analytical) portfolio optimization utility functions, arbitrary cross-correlations and potentially non-Gaussian processes.

Portfolio optimization

- Our optimization universe contains different assets with prices $S_a(t)$ and weights $w_a(t)$ where a is assets index
- The portfolio return is a weighed sum of the asset *returns*, $\Delta X_a = \Delta S_a / S_a$, i.e.

$$\sum_a w_a(t)\,\Delta X_a(t)$$

- To calculate the weights on the next time period we proceed with minimizing/maximizing different utility functions of the distribution parameters of the portfolio increment: the simplest is the **min-var** optimization

Classical Markowitz

The min-var optimization is to find weights w which will minimize the portfolio variance subjected to one constraint, i.e. in vector/matrix notations:

$$\text{minimize } \sigma^2(w) = w^T\, V\, w \quad \text{s.t} \quad w^T\, a = 1$$

The assets *covariance matrix* V elements are often calculated as

$$V_{ab} = \frac{1}{T} \sum_{n=1}^{N_T} \Delta X_a(t_n)\, \Delta X_b(t_n)$$

where the summation run over (business-daily) tenor $\{t_n\}_{n=1}^{N_T}$, s.t. $t - T < t_1 < \cdots < t_{N_T} < t$.
Using a constrained Lagrangian we obtain the following optimal weights and the optimal portfolio variance

$$w^* = \frac{V^{-1}\, a}{a^T\, V^{-1}\, a} \quad \text{and} \quad \sigma^2(w^*) = \frac{1}{a^T\, V^{-1}\, a}$$

Covariance matrix

To simplify the notations we scale the returns

$$X_{a,n} = \Delta X_a(t_n)/\sqrt{\Delta t_n}$$

to obtain

$$V_{ab} = \frac{1}{N_T}\sum_{n=1}^{N_T} X_{a,n}\,X_{b,n}$$

- Of course, the covariance matrix is not necessarily positively defined: either by nature (some assets are linearly dependent) or by calculation errors (MC estimation noise etc)
- However, there are multiple way to regularize it, e.g. using Pastur-Marchenko technique [1], Ledoit-Wolf [3] approach and others, see, for example, [5].
- The number of time-steps N_T and can be rarely above 1000 (4Y), otherwise, the estimated cov matrix will be "outdated".

MC noise for the Markowitz weights.

Let us proceed with our main goal: estimation of the ”MC noise” coming from the covariance matrix summation

$$V_{ab} = \frac{1}{N_T} \sum_{n=1}^{N_T} X_{a,n} X_{b,n}$$

and penetrating into the optimal weights.
Let is decompose the estimated matrix in the exact value (denoted with ”bar”) and the finite-sample noise

$$V = \bar{V} + \Delta V$$

where the noise is Gaussian distribution for large N_T

$$\Delta V_{ab} = \frac{1}{N_T} \sum_{p} (X_{a,n} X_{a,n} - \mathbb{E}[X_a X_b])$$

Here X_a is a theoretical return stochastic variable.

Matrix expansion

Our first technique is based on the matrix expansion for small ΔV.

Example. Let us apply it to the noise of the inverse of the matrix

$$\Delta\left(V^{-1}\right) \equiv V^{-1} - \bar{V}^{-1}$$

Then, ignoring the square of ΔV in the following reasoning

$$\left(\bar{V}^{-1} + \Delta\left(V^{-1}\right)\right)\left(\bar{V} + \Delta V\right) = 1 \;\;\Rightarrow\;\; \Delta\left(V^{-1}\right)\bar{V} + \bar{V}^{-1}\Delta V = 0$$

we obtain the noise of the inverse covariance matrix

$$\Delta\left(V^{-1}\right) \approx -\bar{V}^{-1}\,\Delta V\,\bar{V}^{-1}$$

Comment. As we will see below, the answer for the low number of time-steps (say, corresponding to 1Y) can be sensitive to the second order of ΔV.

Inserting this approximation into the Markowitz formula we obtain[1]

$$w \approx \frac{\left(\bar{V}^{-1} + \Delta\left(V^{-1}\right)\right) a}{a^T \left(\bar{V}^{-1} + \Delta\left(V^{-1}\right)\right) a}$$

Expanding it

$$w \approx \bar{w} + \Delta w$$

around the *exact weights*

$$\bar{w} = \frac{\bar{V}^{-1} \, a}{a^T \, \bar{V}^{-1} \, a}$$

we get the noise of the weights

$$\Delta w \approx -(1 - \bar{w} \, a^T) \, V^{-1} \, \Delta V \, \bar{w}$$

[1]We have removed the star from the weights for brevity.

The noise expectation

We will measure the noise as a matrix expectation

$$\mathbb{E}\left[\Delta w \, \Delta w^T\right]$$

with elements $\mathbb{E}\left[\Delta w_a \, \Delta w_b\right]$. We calculate it as follows:

- The expectations depend on quadratic expression of ΔV
- It can be reduced to a 4-point expectations of X's, i.e. $\mathbb{E}\left[X_n \, X_m \, X_i \, X_j\right]$
- We approximated them supposing that the normalized returns X are Gaussian.

Finally, we obtain a cute expression for the noise matrix

$$\mathbb{E}\left[\Delta w \, \Delta w^T\right] \approx \frac{1}{N_T}\left(\frac{\bar{V}^{-1}}{a^T \, \bar{V}^{-1} \, a} - \bar{w} \, \bar{w}^T\right)$$

The trace of the noise matrix $\mathbb{E}\left[\Delta w \, \Delta w^T\right]$ can be considered as a *one-number measure* of the Markowitz noise

$$\mathcal{N}_M \equiv \mathbb{E}\left[\Delta w^T \Delta w\right] = \operatorname{Tr}\left(\mathbb{E}\left[\Delta w \, \Delta w^T\right]\right)$$
$$\approx \frac{1}{N_T}\left(\frac{\operatorname{Tr}\ \bar{V}^{-1}}{a^T\ \bar{V}^{-1}\ a} - \frac{a^T\ \bar{V}^{-2}\ a}{\left(a^T\ \bar{V}^{-1}\ a\right)^2}\right)$$

Comment. In practice we use use the numerically calculated matrix V instead of its theoretical value $\bar{V}$ in the final formula – this introduces an error of the order $O(N_T^{-2})$ which we can ignore.

Analysis

To qualify the noise, we take an eigenvalues decomposition of the matrix $\bar{V}$ (symmetric positive-definite after the regularization)

$$\bar{V} = U^T \Lambda U$$

where $U^T U = I$ and diagonal matrix Λ contains non-negative eigenvalues. We have

$$\mathrm{Tr}\ \bar{V}^{-1} = \sum_q \lambda_q^{-1} \quad \text{and} \quad a^T \bar{V}^{-k} a = \sum_q b_q^2 \lambda_q^{-k}$$

where $b = U a$ and we sum over the assets $q = 1, \cdots, N_A$.

$$\mathcal{N}_M \approx \frac{1}{N_T} \left(\frac{\sum_q \lambda_q^{-1}}{\sum_q b_q^2 \lambda_q^{-1}} - \frac{\sum_q b_q^2 \lambda_q^{-2}}{\left(\sum_q b_q^2 \lambda_q^{-1}\right)^2} \right)$$

One can easily prove that

$$\mathcal{N}_M \geq 0$$

for any b and λ.

We can consider two special cases:

- The inequality can become equality if one of eigenvalues λ_m^{-1} is dominant. In this case the noise is zero.
- In the opposite case, when the eigenvalues are all equal to each other the noise is maximum, i.e.

$$N_M = \frac{1}{N_T} \frac{N_A - 1}{\sum_q b_q^2}$$

where N_A is the number of assets.

If the portfolio is highly clustered one can come with the HRP optimization [4]. Below we will theoretically prove that it has a smaller MC noise.

HRP or clustered optimization

Let us consider our assets (their returns) forming several *quasi-independent*, clusters or groups:

$$X = \left\{ Y^{(1)}, \cdots, Y^{(H)} \right\}$$

where H is the number of clusters.

Inside each group the correlation is close to one and intra-correlations are close to zero.

1. Calculate the Markowitz weights independently for all clusters

$$w^{(h)} = \frac{{V^{(h)}}^{-1}\, a^{(h)}}{{a^{(h)}}^T\, {V^{(h)}}^{-1}\, a^{(h)}}$$

where the *cluster covariance matrix*

$$V_{ij}^{(h)} = \frac{1}{N_T} \sum_n Y_{i,n}^{(h)}\, Y_{j,n}^{(h)} \quad \text{and} \quad \bar{V}^{(h)} = \mathbb{E}\left[Y^{(h)}\, {Y^{(h)}}^T \right]$$

and the corresponding vectors $a^{(h)}$ taken from the initial ones $a = \left(a^{(1)}, \cdots, a^{(H)} \right)$

The HRP procedure II

2. Calculate a covariance matrix K (H by H) for *clustered* variables

$$C^{(h)} = {w^{(h)}}^T Y^{(h)} \quad \text{for} \quad h = 1, 2$$

defined as

$$K_{hq} = \frac{1}{N_T} \sum_{n,m,p} w_n^{(h)} \; Y_{n,p}^{(h)} \; Y_{m,p}^{(q)} \; w_m^{(q)}$$

It has simplified diagonal elements

$$K_{hh} = \Omega_h^{-1}$$

where we have denoted the quadratic form

$$\Omega_h = {a^{(h)}}^T \; {V^{(h)}}^{-1} \; a^{(h)}$$

The HRP procedure III

3. Calculate the final portfolio with weights ξ_h for the cluster variables

$$\Pi = \xi_1 \, C^{(1)} + \cdots + \xi_H \, C^{(H)}$$

s.t. its variance

$$\sigma^2(\xi) = \mathbb{E}\,[\Pi] = \xi^T K \, \xi$$

is minimized provided that

$$\left(\xi_1 \, w^{(1)}, \cdots, \xi_H \, w^{(H)}\right) \cdot \left(a^{(1)}, \cdots, a^{(H)}\right) = 1$$

This is simply equivalent to

$$\xi_1 + \cdots + \xi_H = \xi \cdot \iota = 1$$

where $\iota = (1, \cdots, 1)$ because $w^{(h)} \cdot a^{(h)} = 1$.

The HRP procedure IV

The optimal values of the clusters weights are given by the Markowitz formula

$$\xi = \frac{K^{-1}\,\iota}{\iota^T\,K^{-1}\,\iota}$$

4. The final portfolio weights, $u^{(h)} = \xi_h\, w^{(h)}$, will be

$$\left(u^{(1)}|\cdots|u^{(H)}\right) = \left(\xi_1\, w_1^{(1)} \cdots \xi_1 w_{N_1}^{(1)}|\cdots|\xi_H w_1^{(H)}, \cdots, \xi_H w_{N_H}^{(H)}\right)$$

The HRP noise

The total portfolio weights noise comes from:

- the cluster weights ξ
- the noise inside the clusters $w^{(h)}$

Summing them up we obtain the final HRP noise

$$\mathcal{N}_C = \frac{1}{N_T}\frac{1}{\Omega}\sum_h \left[\frac{{a^{(h)}}^T \; {V^{(h)}}^{-2} \; a^{(h)}}{\Omega_h} \left(1 - 2\frac{\Omega_h}{\Omega} + 2\frac{\Omega_h}{\Omega}\sum_r \frac{\Omega_r^2}{\Omega^2}\right) \right.$$
$$\left. + \quad \mathrm{Tr}\; {\bar{V}^{(h)}}^{-1} \frac{\Omega_h}{\Omega} \right]$$

where $\Omega = \sum_h \Omega_h$.

The HRP noise is less than the pure Markowitz noise: it is transparent esp. for a large number of clusters when $\Omega_h \ll \Omega$.

Below we confirm the theory with numerical experiments.

Numerical experiments

We set up a clustered correlation matrix with the following clusters on the block diagonal

cluster	0	1	2	3	4	5	6	7	8	9
sizes	10	17	5	17	7	9	15	9	11	3
corrs	0.9	0.8	0.8	0.9	0.8	0.8	0.7	0.8	0.7	0.7

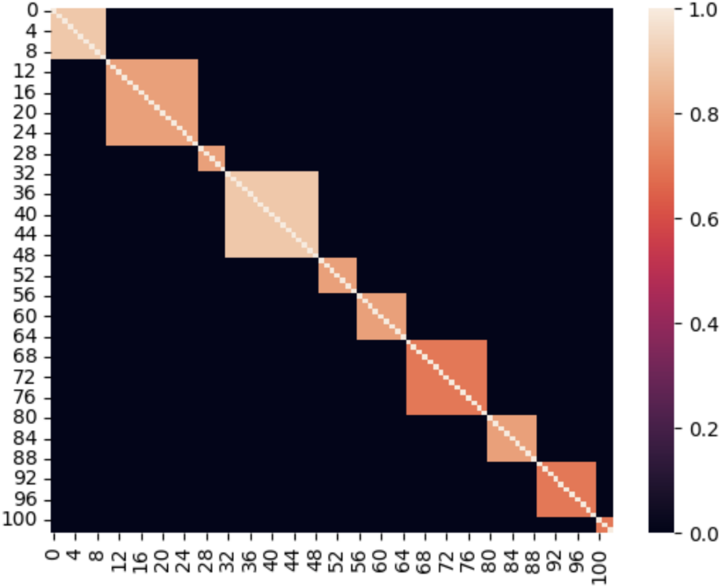
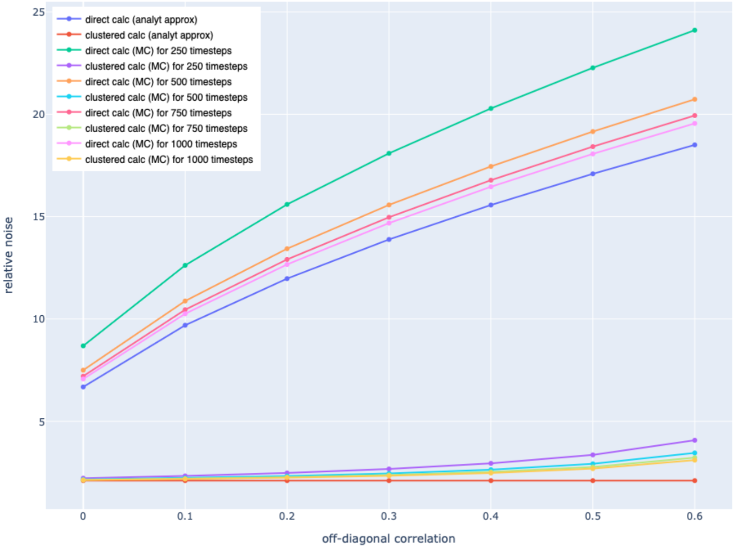


Figure: Clustered (block) correlation matrix.

- The number of assets corresponding to the matrix size is $N_A = 103$
- We perturb the initial correlation matrix with off-cluster values which we vary from 0 (unperturbed) till 0.6
- We simulate N_A Gaussians with these correlation matrices over variable number of time-steps N_T
- We try 250, 500, 750 and 1000 time-steps corresponding approximately to discretized 1Y, 2Y, 3Y and 4Y
- We produce sufficient number of samples for each trajectory – N_A assets over N_T time-steps – to ensure the Monte Carlo convergence
- We plot normalize noise $\sqrt{N_T \, E \Delta w^T \Delta w}$ for the direct Markowitz and $\sqrt{N_T \, \mathbb{E} \left[\Delta u^T \Delta u \right]}$ for the HRP


direct calc (analyt approx)
clustered calc (analyt approx)
direct calc (MC) for 250 timesteps
clustered calc (MC) for 250 timesteps
direct calc (MC) for 500 timesteps
clustered calc (MC) for 500 timesteps
direct calc (MC) for 750 timesteps
clustered calc (MC) for 750 timesteps
direct calc (MC) for 1000 timesteps
clustered calc (MC) for 1000 timesteps
25
20
15
10
5
relative noise
0
0.1
0.2
0.3
0.4
0.5
0.6
off-diagonal correlation

Observations

- A gap between MC noise calculation and the analytics for the Markowitz optimization is due to non-liniear effects. Indeed, in the analytics we have ignored the second order of the covariance matrix noise. Increasing the number of time-steps reduces this gap
- A gap between the HRP MC noise calculation and the analytics is due to the non-linearity *and* the fact that the analytcs ignores the off-diagonal elements
- The impact of the HRP to the noise reduction is significant: 3 times for a pure block structure and 5 for more significant off-diagonal correlations

Conclusions

- We calculated analytical formulas estimating the noise of portfolio optimization weights for both direct Markowitz optimization as well as the HRP one
- Their comparison shows that the HRP is less noisy than the direct Markowitz
- We have confirmed the analytical results by numerical experiments
- One can easily generalize the results for more complicated (but still analytical) portfolio optimizations

References

- Vladimir Marchenko and Leonid Pastur (1967) "Distribution of eigenvalues for some sets of random matrices", *Matematicheskii Sbornik*, 114(4), 507–536.
- Harry Markowitz (1952) "Portfolio Selection", *The Journal of Finance*, 7, 77-91.
- Ledoit, O., and Wolf, M. (2003) "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection", *Journal of Empirical Finance*, 10(5), 603–621
- Marcos López de Prado (2016) "Building Diversified Portfolios that Outperform Out of Sample", *The Journal of Portfolio Management*, 42 (4) 59-69
- Marcos López de Prado (2020), "Machine Learning for Asset Managers", Cambridge Elements in Quantitative Finance