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

Alexandre Antonov, Alexander Lipton and Marcos Lopez de Prado

ADIA, ADIA Lab, Khalifa University

December 4, 2023

Abstract

In this paper we compare two methods of portfolio allocation: the classical Markowitz one and the hierarchical risk parity (HRP) approach. We derive analytical values for the noise of allocation weights coming from the estimated covariance. We demonstrate that the HRP is indeed less noisy (and thus more robust) w.r.t. the classical Markowitz. The second part of the paper is devoted to a detailed analysis of the optimal portfolio variance for which we derive analytical formulas and theoretically demonstrate the superiority of the HRP w.r.t to the Markowitz optimization.

We also address practical outcomes of our analytics. The first one is a fast estimation of the confidence level of the optimization weights calculated for a single (real-life) scenario. The second practical usefulness of the analytics is an HRP portfolio construction criterion which selects assets and clusters minimizing the analytical portfolio variance. We confirm our theoretical results with numerous numerical experiments.

Our calculation technique can be also used in other areas of portfolio optimization.

1 Introduction

One of the most important questions in all of Economics concerns the development of systems that optimally allocate scarce resources. These systems are not unique, and their characteristics adapt to the peculiarity of each scarcity problem. For example, in the crude oil market, buyers arrive at an equilibrium price based on the volumes announced by producers. In contrast, a carmaker determines the amount and price of cars that maximizes its net profit. Investors face a similar question: What is the optimal allocation to various investments, in terms of minimizing the risk of achieving a predefined return? Like with the aforementioned examples, there is not a unique system capable of answering this question in a logical way. In fact, different investors may answer the same question using different methods, in a reflection to their informational sets, biases, or objectives.

In the year 1954, Harry Markowitz proposed a celebrated framework for answering the question of assigning funds to an investable universe [3]. In this framework, investors know the parameters of the multivariate Normal distribution of returns for that investment universe: the true vector of (future) mean returns, and the true (future) covariance matrix of returns. With that information, Markowitz proved that an investor could derive the optimal allocation that would maximize the expected return for a given level of risk, or that it would minimize the level of risk for a given expected return.

It is important to mention, that another pioneer of the minimum variance approach, an Italian mathematician de Finetti, has published his results 14 years before Markowitz in 1940 [1]. Moreover, de Finetti has studied a *constrained* minimization of the variance for positive weights (and given expected returns); see Pressacco & Serafini [7] describing de Finetti's findings under an angle of modern mathematical programming methods.

Needless to say, investors do not know the true vector of future means and the true future covariance matrix of returns. They do not know the sign of the expected means, much less the rounded percentage value. The problem is, Markowitz solutions are notoriously sensitive to even the small changes in these parameters, and this instability increases with the size of the investment universe. Out of the two necessary parameters, the most uncertain is the vector of future means. For this reason, early on many investors opted for estimating the minimum variance portfolio, i.e. the optimal portfolio with minimum risk. This portfolio is convenient because it can be computed without any knowledge of the expected mean returns.

Unfortunately, minimum variance solutions have been met with strong criticism by practitioners. The reason is, it is very difficult to predict the future values of the off-diagonal elements of the covariance matrix. For example, the correlation between stocks and bonds may flip unpredictably from negative to positive, however the volatility of stocks is relatively stable over time. Markowitz's minimum variance portfolios are very sensitive to changes in correlations, which makes its solutions not robust. To address this concern, in the 1990s practitioners proposed so-called risk parity approaches. The general idea is, within the covariance matrix, investors are more confident about the main diagonal than about the off-diagonal elements, hence the allocation should be informed by variances rather than covariances.

One criticism of this risk parity approach is that it entirely throws out all correlation information. Surely, we may not be able to predict the correlation between stocks and bonds, but we may be able to predict the correlation between two stocks in the same sector, or in the same region, or stocks in the same supply-chain, etc. In 2016, Lopez de Prado proposed the hierarchical risk parity (HRP) approach [4] as a compromise between the two radical approaches of Markowitz's minimum variance portfolio (which assumes perfect knowledge of the future covariance matrix) and risk parity (which assumes perfect ignorance of all correlations). In his seminal paper, Lopez de Prado showed vis Monte Carlo experiments that "HRP delivers lower out-of-sample variance than [Markowitz's minimum variance portfolio], even though minimum-variance is Markowitz's optimization objective. HRP also produces less risky portfolios out-of-sample compared to traditional risk parity methods." In this paper, we reach the same conclusions through analytical methods.

Namely, under Gaussian assumptions of the asset time-series, we derive analytical approximation of a noise of the allocation weights coming from the estimated covariance.

Our method, based on an expansion of a *noise* of the covariance matrix, is applicable when the number of assets is *moderate* w.r.t. the the sample size used to estimate the covariance. This is in contrast with the Marchenko-Pastur criterion [2] where the number of assets is supposed to be *comparable* with the sample size.

Natural noise measures – such as expected variance of the allocation weights, its trace as well as the optimal portfolio risk variance – can be calculated analytically for the Markowitz optimization. The resulting formulas are quite compact and can be easily implemented.

To treat the HRP case we assume that the asset clusters are already detected (see, for example, [6]) and that the intra-cluster correlations are relatively low. This permits to derive the noise measures for the HRP case: the formulas are only slightly more complicated than these for the Markowitz optimization. We demonstrate that the HRP is indeed less noisy (and thus more robust) w.r.t. the classical Markowitz.

Another important part of the paper is devoted to a detailed analysis of the optimal *portfolio variance* for both Markowitz and HRP methods. We derive the portfolio variance analytical formulas and theoretically demonstrate that the out-of-sample HRP is indeed less noisy and more robust that the Markowitz optimization.

The first *practical* outcome of our analytics can be a fast estimation of the *confidence level* of the optimization weights calculated for a single (real-life) scenario. The second practical usefulness of the noise analytics can be a portfolio selection for which minimizes the analytical HRP portfolio variance.

We confirm the theoretical results using multiple numerical experiments based on Monte Carlo simulations. The focus is made on the weights noise and the optimal portfolio statistics for in- and out-of-sample cases.

Finally, notice that we derive the formulas for the min-variance optimization, Gaussian 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.

The paper is organized as follows. In the main body of the paper we announce the main results with brief descriptions of the derivation logic, s.t. all (relatively tedious) calculations are put in the appendixes. In Section 2 we calculate the Markowitz optimization noise measures and comment on the formulas applicability criterion. In Section 3 we remind the HRP method work-flow, derive the noise measures and compare the results with the Markowitz base line. Next, in Section 4 we provide a detailed analysis of the optimal portfolio variance statistics for the Markowitz and the HRP for both in- and out-of-sample cases. We present numerical experiments in Section 5.

2 The Markowitz optimization and its noise

Our optimization universe contains different assets with returns $X_i(t)$ and weights $w_i(t)$ where i is an asset index. The portfolio return is a weighed sum of the asset returns i.e. $\sum_i w_i(t)\, X_i(t)$. To calculate the weights on the next time period we proceed with optimizing different utility functions of the distribution parameters of the portfolio increment: the simplest procedure is the **min-var** optimization.

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

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

The assets *covariance matrix* V elements are often estimated from returns time-series

$$V_{ij} = \frac{1}{N_T} \sum_{n=1}^{N_T} X_{i,n}\, X_{j,n} \tag{2}$$

where the summation run over (business-daily) dates $\{t_n\}_{n=1}^{N_T}$ and $X_{i,n} = X_i(t_n)$. The asset indices i, j go from 1 to N_A.
Using a constrained Lagrangian we obtain the following optimal weights

$$w^* = \frac{V^{-1}\, a}{a^T\, V^{-1}\, a} \tag{3}$$

with the corresponding optimal variance

$$\sigma^2(w^*) = \frac{1}{a^T\, V^{-1}\, a} \tag{4}$$

Of course, the covariance matrix is not necessarily positively defined: either by nature (some assets are linearly dependent) or by calculation errors (Monte Carlo estimation noise etc). However, for our calculations we suppose that, thanks to its clustered structure, the covariance matrix is invertable.

The sample size N_T can be rarely above five years of the daily data, otherwise, the estimated covariance matrix will be "stalled". In general, the number of *assets* N_A can be either small w.r.t. N_T or comparable to it. In both cases, the exact values are blurred by the noise from their *exact* positions corresponding to $N_T \to \infty$. In this paper we concentrate on **moderate number of assets**. Theoretically, it means that $N_A/N_T \ll 1$ but, in practice, this coefficient can be large enough, say, start with 1/2 or 1/3, to attain a reasonable accuracy. In the end of this Section we address the applicability criterion in more details.

Monte Carlo noise for the allocation weights. Let us proceed with our main goal: estimation of the "Monte Carlo noise" coming from the covariance matrix summation (2) and penetrating into the optimal weights.

Let us decompose the estimated matrix in the exact value (denoted with "bar") and the finite-sample noise

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

where the noise has a Gaussian distribution for large N_T

$$\Delta V_{ij} = \frac{1}{N_T} \sum_{n=1}^{N_T} (X_{i,n}\, X_{j,n} - \mathbb{E}\,[X_i\, X_j])$$

Here X_i is a theoretical return *stochastic variable.*

Our first calculation tool is the matrix expansion for small ΔV. For example, let us apply it to the noise of the inverse of the matrix

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

Ignoring the square of ΔV in the following reasoning

$$\left(\bar{V}^{-1} + \Delta\left(V^{-1}\right)\right)\, (\bar{V} + \Delta V) = 1 \quad \Rightarrow \quad \Delta\left(V^{-1}\right)\, \bar{V} + \bar{V}^{-1}\, \Delta V \approx 0 \tag{5}$$

we obtain the noise of the inverse covariance matrix

$$\Delta\left(V^{-1}\right) \approx -\bar{V}^{-1}\, \Delta V\, \bar{V}^{-1} \tag{6}$$

As we will see below, the answer for the small sample size (say, corresponding to one year of daily data) can be sensitive to the second order of ΔV but this dependence will radically decrease for three or four years interval.

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

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

Expanding it

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

around the *exact weights*

$$\bar{w} = \frac{\bar{V}^{-1}\, a}{a^T\, \bar{V}^{-1}\, a} \tag{7}$$

we get the noise of the weights

$$\Delta w \approx -(I - \bar{w}\, a^T)\, V^{-1}\, \Delta V\, \bar{w} \tag{8}$$

The most natural noise measure is the covariance

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

with elements $\mathbb{E}\left[\Delta w_i\, \Delta w_j\right]$. To estimate it we notice that the expectations in hand depends on quadratic expression of ΔV, namely, on $\mathbb{E}\left[\Delta V\, \bar{w}\, \bar{w}^T\, \Delta V\right]$. In Appendix A we prove a general result for an arbitrary matrix M

$$\mathbb{E}\left[\Delta V\, M\, \Delta V\right] = \frac{1}{N_T}\left(\mathbb{E}\left[X\, X^T\, (X^T\, M\, X)\right] - \bar{V}\, M\, \bar{V}\right)$$

which gives the desired expectation for $M = \bar{w}\, \bar{w}^T$. This expression depends on 4-point average of X's, i.e. $\mathbb{E}\left[X_n\, X_m\, X_i\, X_j\right]$. We can exactly evaluate them assuming that the normalized returns X are Gaussian which leads to the following general relationship

$$\mathbb{E}\left[\Delta V\, M\, \Delta V\right] = \frac{1}{N_T}\left(\bar{V}\, \mathrm{Tr}(\bar{V}\, M) + \bar{V}\, M^T\, \bar{V}\right) \tag{9}$$

which permits us obtain an elegant expression for the noise matrix

$$\mathbb{E}\left[\Delta w\, \Delta w^T\right] \approx \frac{1}{N_T}\left(\frac{\bar{V}^{-1}}{a^T\, \bar{V}^{-1}\, a} - \bar{w}\, \bar{w}^T\right) \tag{10}$$

The trace of the noise matrix expectation can be considered as a *one-number measure* of the Markowitz noise such that

$$\mathcal{N}_M \equiv \mathbb{E}\left[\Delta w^T \Delta w\right] = \mathrm{Tr}\left(\mathbb{E}\left[\Delta w\, \Delta w^T\right]\right) \approx \frac{1}{N_T}\left(\frac{\mathrm{Tr}\ \bar{V}^{-1}}{a^T\, \bar{V}^{-1}\, a} - \frac{a^T\, \bar{V}^{-2}\, a}{\left(a^T\, \bar{V}^{-1}\, a\right)^2}\right) \tag{11}$$

Its slight generalization for some matrix M will be used below for the portfolio variance studies

$$\mathbb{E}\left[\Delta w^T\, M\, \Delta w\right] \approx \frac{1}{N_T}\left(\frac{\mathrm{Tr}\ \left(\bar{V}^{-1}\, M\right)}{a^T\, \bar{V}^{-1}\, a} - \bar{w}^T\, M\, \bar{w}\right) \tag{12}$$

Of course, in practice, we use the *estimated* matrix V instead of its theoretical value in our formulas (10-11).

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

Noise inequality. To demonstrate a non-negativity of the noise expectation (11) we proceed as follows. Denote the eigenvalue decomposition of the exact covariance matrix as

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

with eigenvalues $\bar{\Lambda}_{ij} = \delta_{ij}\bar{\lambda}^{(i)}$ and eigenvectors matrix $\bar{U} = \left\{\bar{u}^{(1)}, \cdots, \bar{u}^{(N_A)}\right\}$,

$$\bar{V}\,\bar{u}^{(i)} = \bar{\lambda}^{(i)}\,\bar{u}^{(i)}$$

Thus, we rewrite the components of (11) as

$$\mathrm{Tr}\;\bar{V}^{-1} = \sum_n \bar{\lambda}_n^{-1} \quad \text{and} \quad a^T\,\bar{V}^{-k}\,a = \sum_n \bar{b}_n^2\,\bar{\lambda}_n^{-k}$$

for $\bar{b} = \bar{U}\,a$. Clearly,

$$\mathrm{Tr}\;\bar{V}^{-1}\;a^T\,\bar{V}^{-1}\,a \geq a^T\,\bar{V}^{-2}\,a \tag{13}$$

due to

$$\sum_m \bar{\lambda}_m^{-1}\sum_n \bar{b}_n^2\,\bar{\lambda}_n^{-1} \geq \sum_n \bar{b}_n^2\,\lambda_n^{-2}$$

because

$$\sum_m \bar{\lambda}_m^{-1}\sum_n \bar{b}_n^2\,\bar{\lambda}_n^{-1} - \sum_k \bar{b}_k^2\,\bar{\lambda}_k^{-2} = \sum_{m\neq n} \bar{\lambda}_m^{-1}\bar{b}_n^2\,\bar{\lambda}_n^{-1} \geq 0$$

Indeed, all the elements in the last summation are positive. Thus, after the diagonalization, the Markowitz noise (11) looks as follows

$$\mathcal{N}_M \approx \frac{1}{N_T}\frac{\sum_{m\neq n}\bar{\lambda}_n^{-1}\,\bar{\lambda}_m^{-1}\bar{b}_n^2}{\left(\sum_n \bar{b}_n^2\,\bar{\lambda}_n^{-1}\right)^2}$$

Applicability criterion. To assess the validity of our noise formula we return to the inverse covariance matrix noise calculation in (5) and notice that in the approximation we ignored the following quadratic term

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

If this quadratic term is small *in average*, our first-order expansion is valid. Using the general expectation formula (9) we readily obtain

$$\mathbb{E}\left[\Delta\left(V^{-1}\right)\,\Delta V\right] = -\frac{N_A+1}{N_T}\,I$$

which gives us *the criterion of moderate number of assets*

$$\frac{N_A}{N_T} \ll 1$$

Note that below, while studying the portfolio variance, we will also go *beyond* the leading order in N_A/N_T and reach a superior approximation quality.

3 HRP or Clustered Optimization

In [4] it was demonstrated that a clusterization can help with the noise reduction. Consider our assets (their returns) forming several *quasi-independent* groups or clusters

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

We denote the number of assets inside cluster h as N_h (they sum up into the total number of assets $\sum_{h=1}^{H} N_h = N_A$).

Inside each cluster the correlation is large while intra-cluster correlations are close to zero. The clustered optimization procedure starts with applying the Markowitz optimization *independently* for each cluster. This determines an optimal sub-portfolio allocation inside each cluster. The next step is to form a portfolio consisting of cluster sub-portfolios as assets. Finally, using the Markowitz optimization for this portfolio of the sub-portfolios, we come up with the final allocation: the resulting asset weights are the optimal sub-portfolio weights times the asset weight inside each sub-portfolio.

The formal steps are:

1. Calculate the Markowitz weights $w^{(h)}$ independently for each cluster $h = 1, \cdots, H$ using (3)

$$w^{(h)} = \frac{{V^{(h)}}^{-1}\, a^{(h)}}{{a^{(h)}}^T\, {V^{(h)}}^{-1}\, a^{(h)}} \tag{14}$$

where the cluster covariance matrix is estimated as

$$V_{ij}^{(h)} = \frac{1}{N_T} \sum_{n=1}^{N_T} Y_{i,n}^{(h)}\, Y_{j,n}^{(h)} \tag{15}$$

with the corresponding normalizers $a^{(h)}$ taken from the initial ones $a = \left(a^{(1)}, \cdots, a^{(H)}\right)$. We also denote the theoretical (infinite sample) covariance matrix as

$$\bar{V}^{(h)} = \mathbb{E}\left[Y^{(h)}\, {Y^{(h)}}^T\right] \tag{16}$$

2. Calculate a covariance matrix K (H by H) for *clustered* variables (cluster sub-portfolios)

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

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 due to (4)

$$K_{hh} = {w^{(h)}}^T\, V^{(h)}\, w^{(h)} = \frac{1}{\Omega_h}$$

where we have denoted the inverse cluster risk as Ω_h. Using (14) it can be shown that

$$\Omega_h = {a^{(h)}}^T\, {V^{(h)}}^{-1}\, a^{(h)} \tag{17}$$

3. Calculate the cluster weights ξ_h for the cluster variables $C^{(h)}$

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

For this we minimize the portfolio Π variance

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

provided that

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

Note that this normalization condition 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 optimal values of the clusters weights are given by the Markowitz formula (3)

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

4. Determine the final portfolio full weights, $u^{(h)} = \xi_h\, w^{(h)}$,

$$\left(u^{(1)}|\cdots|u^{(H)}\right) = \left(\xi_1\, w_1^{(1)} \cdots \xi_1 w_{N_1}^{(1)}|\cdots|\xi_H w_1^{(H)}, \cdots, \xi_H w_{N_H}^{(H)}\right) \tag{18}$$

As in the Markowitz case, we denote the *theoretical* HRP components with the bar symbol, e.g. $\bar{u}$.

The total portfolio weights noise comes from the cluster weights ξ_h as well as from the assets weights inside the clusters $w^{(h)}$. To simplify calculations, we separate the noise coming from the diagonal blocks of the covariance matrix $V^{(h)}$ and the *off-diagonal* ones

$$\delta V_{ij}^{(h,q)} = \frac{1}{N_T} \sum_n Y_{i,n}^{(h)}\, Y_{j,n}^{(q)} \quad \text{for} \quad h \neq q$$

We put a small delta in front of the off-diagonal covariance matrix because its average value is zero (or small enough) by the assumption. The second reason for the small delta notation is to distinguish the off-diagonal noise from the block-diagonal noise

$$\Delta V_{ij}^{(h)} = \frac{1}{N_T} \sum_n \left(Y_{i,n}^{(h)} \, Y_{j,n}^{(h)} - \mathbb{E}\left[Y_i^{(h)} \, Y_j^{(h)} \right] \right)$$

denoted with a *capital* delta. These noises can be separated because their product expectations are zero

$$\mathbb{E}\left[\delta V_{ij}^{(h,q)} \, \Delta V_{i'j'}^{(h')} \right] = 0$$

for all cluster/element indexes due to zero correlations between different clusters elements $\mathbb{E}\left[Y_i^{(h)} \, Y_{i'}^{(h')} \right] = 0$.

As we will see below *the optimal weights for the Markowitz and HRP methods are identical if the covariance matrix is a block-diagonal one, i.e. cross-cluster correlations are strictly zero*. This means that the noise difference between the Markowitz and the HRP comes from *cross-cluster correlations*. In Appendix B we prove the following analytical formula for the HRP expected noise

$$\begin{aligned} \mathcal{N}_C &= \mathbb{E}\left[\Delta u^T \, \Delta u \right] = \sum_h \mathbb{E}\left[\Delta u^{(h)T} \, \Delta u^{(h)} \right] \\ &\simeq \frac{1}{N_T} \frac{1}{\bar{\Omega}} \left(\sum_h \operatorname{tr}\left(\bar{V}^{(h)-1} \right) \frac{\bar{\Omega}_h}{\bar{\Omega}} + \sum_h \frac{a^{(h)T} \, \bar{V}^{(h)-2} \, a^{(h)}}{\bar{\Omega}_h} \left(1 - 2 \frac{\bar{\Omega}_h}{\bar{\Omega}} \right) \right) \end{aligned} \tag{19}$$

where $\bar{V}^{(h)}$ is an *exact* covariance matrix of h-th cluster (16). Also, we have defined the exact inverse cluster covariance as $\bar{\Omega}_h$ and $\bar{\Omega} = \sum_h \bar{\Omega}_h$. As we have mentioned in Section 2, for practical noise calculation we use the estimated matrix instead of its theoretical value. This introduces an error of the order $O(N_T^{-2})$ which we can ignore.

A more general formula which we will use in the portfolio risk calculation is a simple modification of the formula (19)

$$\begin{aligned} &\mathbb{E}\left[\Delta u^T \, M_B \, \Delta u \right] = \sum_h \mathbb{E}\left[\Delta u^{(h)T} \, M^{(h)} \, \Delta u^{(h)} \right] \\ &\simeq \frac{1}{N_T} \frac{1}{\bar{\Omega}} \left(\sum_h \operatorname{tr}\left(\bar{V}^{(h)-1} \, M^{(h)} \right) \frac{\bar{\Omega}_h}{\bar{\Omega}} + \sum_h \frac{a^{(h)T} \, \bar{V}^{(h)-1} \, M^{(h)} \, \bar{V}^{(h)-1} \, a^{(h)}}{\bar{\Omega}_h} \left(1 - 2 \frac{\bar{\Omega}_h}{\bar{\Omega}} \right) \right) \end{aligned} \tag{20}$$

where M_B is a block matrix which the same dimensions as the block variance matrix.

One can easily demonstrate that the HRP noise (19) is always less than the direct Markowitz one (11) using arguments similar to the previous section ones. Indeed, under our assumption of zero intra cluster correlations, the exact covariance matrix $\bar{V}$ is a block-one, such that its inversion simply consists on inversions of the cluster covariance matrices $\bar{V}^{(h)}$. It is easy to see that the Markowitz expected noise (11) can be written as

$$\mathcal{N}_M \simeq \frac{1}{N_T} \frac{1}{\bar{\Omega}} \left(\sum_h \operatorname{Tr} \bar{V}^{(h)-1} - \frac{\sum_h a^{(h)T} \, \bar{V}^{(h)-2} \, a^{(h)}}{\bar{\Omega}} \right) \tag{21}$$

This permits us to prove that the the difference $N_M - N_C$ is always non-negative

$$\mathcal{N}_M - \mathcal{N}_C \geq \frac{1}{N_T} \frac{1}{\bar{\Omega}^2} \sum_h \left(\operatorname{Tr} \bar{V}^{(h)-1} - \frac{a^{(h)T} \, \bar{V}^{(h)-2} \, a^{(h)}}{\bar{\Omega}_h} \right) \left(\bar{\Omega} - \bar{\Omega}_h \right)$$

Indeed, the non-negativity of the first multiplier was proved in (13) and $\bar{\Omega} = \sum_q \bar{\Omega}_q \geq \bar{\Omega}_h$ because all inverse cluster risks are non-negative ($\bar{\Omega}_h \geq 0$) due to (17).

In the next section we address the main practical application of our theory: the Markowitz and the HRP portfolio risk calculations.

4 Portfolio variance statistics

In this Section we will evaluate the portfolio variance (risk) statistics: its expectation and standard deviation with respect to movements of the covariance matrix. We will consider two important cases: in-sample (IS) and out-of-sample (OOS).

The IS portfolio variance is simply

$$\sigma^2 = w^T V w \tag{22}$$

where the weights are constructed using the *portfolio* covariance matrix V as in (3).

The OOS case risk

$$\tilde{\sigma}^2 = w^T \tilde{V} w \tag{23}$$

is when the *portfolio* covariance matrix $\tilde{V}$ is independent of the *weights* covariance matrix V. This is obviously the real-life case when we apply historically calculated weights to future returns which form a future covariance matrix. Indeed, the estimated variance reads

$$\hat{\sigma}^2 = \frac{1}{N_T} \sum_{n=1}^{N_T} \left(\sum_i w_i X_{i,n} \right)^2$$

where the optimal weights w are calculated at t_0 by the Markowitz formula (3) with the covariance matrix estimated by (2) with returns $X_{i,-N_T}, \cdots, X_{i,-1}$. Thus, the OOS variance can be rewritten in the form of (23) where the weights covariance matrix is

$$V_{ij} = \frac{1}{N_T} \sum_{n=-N_T}^{-1} X_{i,n} X_{j,n}$$

while the portfolio covariance matrix is

$$\tilde{V}_{ij} = \frac{1}{N_T} \sum_{n=1}^{N_T} X_{i,n} X_{j,n}$$

Obviously, the weights matrix increment ΔV is independent on the portfolio one $\Delta \tilde{V}$.

While working with the portfolio variance it is important to go *beyond* the leading order in the number of samples $1/N_T$. Indeed, as we will see below, the IS and OOS portfolio risks coincide in a limit of a large number of samples but the second order effect is quite sizeable for standard portfolios.

Let us pass now to the IS and OSS portfolio variances.

Portfolio variances. A optimal portfolio IS variance is given by (22). For the Markowitz weights (3) the risk is minimal (4). To estimate its statistics, we can proceed directly using perturbation of the variance in its denominator but, instead, we will take another way which is more intuitive and explanatory. Indeed, the noise of the risk comes through the optimal w and V, i.e.

$$\begin{aligned}
\sigma^2 &= (\bar{w} + \Delta w)^T (\bar{V} + \Delta V)(\bar{w} + \Delta w) \\
&= \bar{w}^T \bar{V} \bar{w} + \underbrace{\bar{w}^T \Delta V \bar{w} + 2\bar{w}^T \bar{V} \Delta w}_{\text{1st order}} + \underbrace{2\bar{w}^T \Delta V \Delta w + \Delta w^T \bar{V} \Delta w}_{\text{2nd order}} + \underbrace{\Delta w^T \Delta V \Delta w}_{\text{3rd order}}
\end{aligned}$$

A contribution of the weights noise in the first order $\bar{w}^T \bar{V} \Delta w$ cancels out. Due to the normalization constraint for both exact and realized weight, $a^T \bar{w} = 1$ and $a^T w = 1$, the weights noise is perpendicular to a, i.e. $a^T \Delta w = 0$. As far as $\bar{w}^T \bar{V}$ is proportional to a, $\bar{w}^T \bar{V} \Delta w = 0$, giving

$$\sigma^2 = \bar{\sigma}^2 + \bar{w}^T \Delta V \bar{w} + 2\bar{w}^T \Delta V \Delta w + \Delta w^T \bar{V} \Delta w + \Delta w^T \Delta V \Delta w \tag{24}$$

where we have denoted the theoretical (infinite number of samples) risk as

$$\bar{\sigma}^2 = \bar{w}^T \bar{V} \bar{w} \tag{25}$$

Similarly, the OOS risk (23) can be expanded as

$$\tilde{\sigma}^2 = \bar{\sigma}^2 + \bar{w}^T \Delta \tilde{V} \bar{w} + 2\bar{w}^T \Delta \tilde{V} \Delta w + \Delta w^T \bar{V} \Delta w + \Delta w^T \Delta \tilde{V} \Delta w \tag{26}$$

Now let us start with the risk expectation calculations followed by the risk standard deviation.

The risk expectation. In these studies we go beyond the leading order in $1/N_T$ to explain the effect of the risk noise increase when we switch from the in-sample to out-of-sample portfolio. Taking expectation of (24)

$$\mathbb{E}\left[\sigma^2\right] = \bar{\sigma}^2 + 2\mathbb{E}\left[\bar{w}^T \Delta V \Delta w\right] + \mathbb{E}\left[\Delta w^T \bar{V} \Delta w\right] + O\left(\frac{1}{N_T^2}\right)$$

we obtain

$$\mathbb{E}\left[\sigma^2\right] = \bar{\sigma}^2 - \mathbb{E}\left[\Delta w^T \bar{V} \Delta w\right] + O\left(\frac{1}{N_T^2}\right) \tag{27}$$

Here we have used the identity[2]

$$\mathbb{E}\left[\bar{w}^T \Delta V \Delta w\right] = -\mathbb{E}\left[\Delta w^T \bar{V} \Delta w\right] + O\left(\frac{1}{N_T^2}\right)$$

In the OOS case, the risk expectation is simpler due to the independence on the weights and portfolio noises, ΔV and $\Delta \tilde{V}$, resulting in $\mathbb{E}\left[\bar{w}^T \Delta \tilde{V} \Delta w\right] = 0$. This leads to

$$\mathbb{E}\left[\tilde{\sigma}^2\right] = \bar{\sigma}^2 + \mathbb{E}\left[\Delta w^T \bar{V} \Delta w\right] + O\left(\frac{1}{N_T^2}\right) \tag{28}$$

We see that the IS risk expectation is always *smaller* than the OSS one by $2\mathbb{E}\left[\Delta w^T \bar{V} \Delta w\right]$. The raison is obvious: the risk is *explicitly* minimized in the IS case while the OOS risk is not.

Now we will evaluate the IS and OOS risks expectations for the Markowitz and the HRP methods. Thanks to the simplified expressions (27-28) this calculation can be easily performed.

For the Markowitz case we use the expected noise formula (12) for the underlying expectation

$$\mathbb{E}\left[\Delta w^T \bar{V} \Delta w\right] = \frac{N_A - 1}{N_T}\bar{\sigma}^2 + O\left(\frac{1}{N_T^2}\right) \tag{29}$$

to obtain

$$\begin{aligned} \mathbb{E}\left[\sigma_M^2\right] = &\quad \bar{\sigma}^2 \left(1 - \frac{N_A - 1}{N_T}\right) + O\left(\frac{1}{N_T^2}\right) \\ \mathbb{E}\left[\tilde{\sigma}_M^2\right] = &\quad \bar{\sigma}^2 \left(1 + \frac{N_A - 1}{N_T}\right) + O\left(\frac{1}{N_T^2}\right) \end{aligned} \tag{30}$$

where we have put a subscript M to emphasize that the risk belongs to the Markowitz portfolio.

For the HRP case we proceed in the similar manner calculating the risk expectations for the pure block case such that the theoretical matrix $\bar{V}$ is the block one, i.e. $\bar{V} = \bar{V}_B$. The expectation underlying the formulas (27-28) can be calculated using the generalized noise (20)

$$\mathbb{E}\left[\Delta u^T \bar{V}_B \Delta u\right] = \frac{1}{N_T}\frac{1}{\bar{\Omega}}\left(\sum_h N_h \frac{\bar{\Omega}_h}{\bar{\Omega}} + \sum_h \left(1 - 2\frac{\bar{\Omega}_h}{\bar{\Omega}}\right)\right) \tag{31}$$

where N_h is the number of assets in a cluster h. This leads to the final HRP answer for both IS and OSS cases

$$\begin{aligned} \mathbb{E}\left[\sigma_C^2\right] = &\quad \bar{\sigma}^2 \left(1 - \frac{H - 1 + \sum_h (N_h - 1)\frac{\bar{\Omega}_h}{\bar{\Omega}}}{N_T}\right) + O\left(\frac{1}{N_T^2}\right) \\ \mathbb{E}\left[\tilde{\sigma}_C^2\right] = &\quad \bar{\sigma}^2 \left(1 + \frac{H - 1 + \sum_h (N_h - 1)\frac{\bar{\Omega}_h}{\bar{\Omega}}}{N_T}\right) + O\left(\frac{1}{N_T^2}\right) \end{aligned} \tag{32}$$

[2] It follows from

$$\mathbb{E}\left[\bar{w}^T \Delta V \Delta w\right] + \mathbb{E}\left[\Delta w^T \bar{V} \Delta w\right] = \mathbb{E}\left[\Delta(w^T V)\Delta w\right] + O\left(\frac{1}{N_T^2}\right) = O\left(\frac{1}{N_T^2}\right)$$

valid due to a proportionality of $\Delta(w^T V)$ to the normalization vector a and the constrain $a^T \Delta w = 0$.

where we put the subscript C to address the *clustered* (HRP) risk.

It is easy to see that the HRP risk expectation first order correction (denoted as δ_C) is always less than the Markowitz one δ_M, i.e.

$$\delta_C = \frac{H - 1 + \sum_h (N_h - 1) \frac{\bar{\Omega}_h}{\bar{\Omega}}}{N_T} \leq \frac{N_A - 1}{N_T} = \delta_M$$

Indeed, this inequality is equivalent to an obvious inequality

$$\sum_h (N_h - 1) \left(1 - \frac{\bar{\Omega}_h}{\bar{\Omega}}\right) \geq 0$$

Having the analytical expression of the HRP correction we can chose the portfolio/cluster composition to minimize the risk. For example, if the clusters contains the same number of elements, the minimal HRP correction corresponds to the number of clusters around $\sqrt{N_A}$, s.t. its minimal value

$$\min \delta_C \simeq 2 \frac{\sqrt{N_A}}{N_T} \tag{33}$$

can be much less than the Markowitz one

$$\delta_M \simeq \frac{N_A}{N_T} \tag{34}$$

We see that that the expected risk for our 4 cases: IS/OOS and Markowitz/HRP satisfies the following inequality:

$$\mathbb{E}\left[\sigma_M^2\right] \leq \mathbb{E}\left[\sigma_C^2\right] \leq \bar{\sigma}^2 \leq \mathbb{E}\left[\tilde{\sigma}_C^2\right] \leq \mathbb{E}\left[\tilde{\sigma}_M^2\right] \tag{35}$$

The IS Markowitz risk is less than the IS HRP because the former directly minimizes the risk. On the other hand, for the OOS cases the HRP risk expectation is smaller than the Markowitz one because the HRP weights are less noisy than the Markowitz ones. This is also important for other aspects of the portfolio *robustness*: a smaller weights noise leads to a lower turnover and transaction costs and makes the optimization less sensitive to sudden market changes.

The risk variance. The IS portfolio variance (risk) variance can be easily evaluated in the leading order

$$\mathbb{V}\left[\sigma^2\right] = \frac{2}{N_T}\, \bar{\sigma}^4 + O\left(\frac{1}{N_T^2}\right) \tag{36}$$

For this we have selected the first order term $\bar{w}^T \Delta V \bar{w}$ from the risk expansion (24) and calculated its square expectation using the general formula (11).

The same formula is valid for the OOS case

$$\mathbb{V}\left[\tilde{\sigma}^2\right] = \frac{2}{N_T}\, \bar{\sigma}^4 + O\left(\frac{1}{N_T^2}\right) \tag{37}$$

As far as the *theoretical* risks for the Markowitz and the HRP are identical (due to $\bar{u} = \bar{w}$) their *numerical* risk values coincide in the leading order. Going beyond the first order is *much* more complicated because the expectations in hand contain averages of the fourth order in ΔV. That is why we can repeat the qualitative arguments of the previous paragraph and come up with the following inequality *in the higher order*

$$\mathbb{V}\left[\sigma_M^2\right] \leq \mathbb{V}\left[\sigma_C^2\right] \leq \frac{2}{N_T}\, \bar{\sigma}^4 \leq \mathbb{V}\left[\tilde{\sigma}_C^2\right] \leq \mathbb{V}\left[\tilde{\sigma}_M^2\right] \tag{38}$$

We will observe this equality in next Section of numerical experiments.

Finally, let us notice that for the Markowitz case, going beyond the leading order in the *risk variance* is less important for practical applications than for the *risk expectation*. Indeed, comparing corrections to the risk expectations (34) with its normalized theoretical standard deviation, i.e.

$$\frac{N_A}{N_T} \text{ v.s. } \sqrt{\frac{2}{N_T}}$$

we conclude that the expectation corrections will dominate the risk standard deviation if the number of assets is more than a square root of the number of samples. This means that the difference between the Markowitz portfolio risk and the HRP one for **a single scenario** is mostly described by a difference between their **expected** values rather than by **standard deviations** of the risk.

5 Numerical Experiments

For 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

Table 1: Correlation matrix cluster composition.

The corresponding number of assets is $N_A = 103$. The matrix can be visualized below

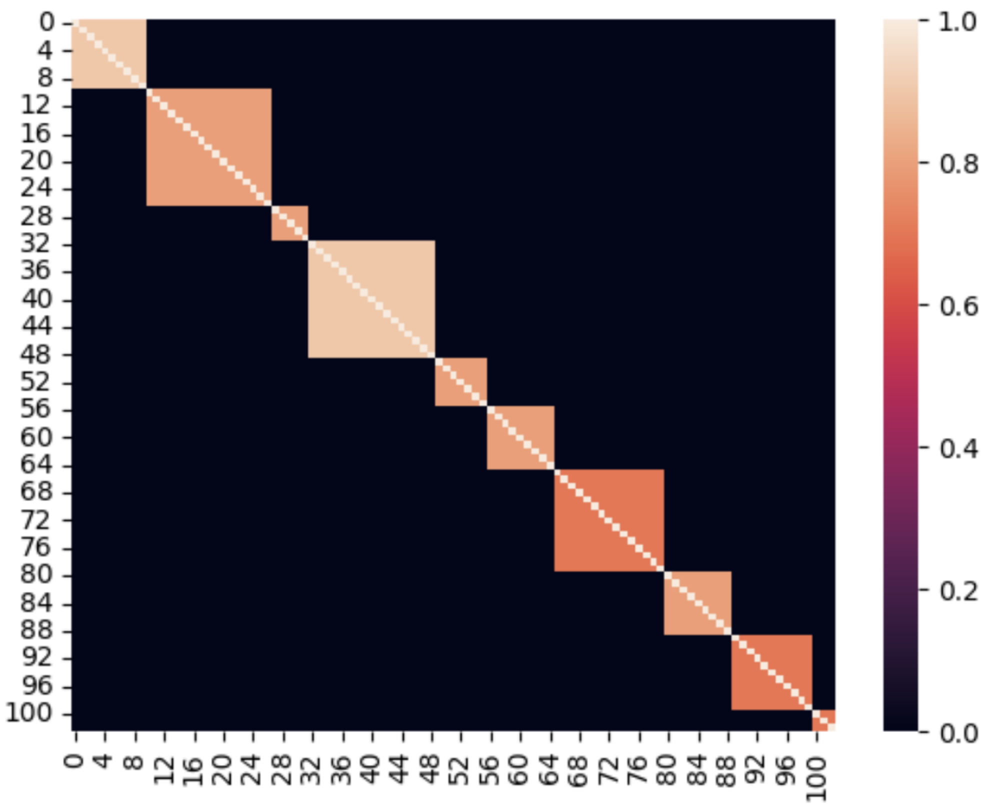

Figure 1: Clustered (block) correlation matrix.

In our experiments we perturb the initial correlation matrix with *off-cluster* (or off-block) values which we vary from 0 (unperturbed) till 60%. Then, we simulate N_A Gaussians with these correlation matrices[3] over variable number of samples N_T: we try 250, 500, 750 and 1000 time-steps corresponding approximately to 1, 2, 3 and 4 years of daily data. We produce 10,000 of such Monte Carlo trajectories – N_A assets over N_T samples () – to ensure the Monte Carlo convergence. In the first group of experiments we demonstrate the analytics validity and quantify the noise reduction of the HRP w.r.t. the Markowitz.

Noise measurements. We start with experiments for 500 timesteps (2 years of daily data) and zero off-block correlation. We output analytical (exact) values $\bar{w}$ as well as these for a typical simulation scenario for both Markowitz and HRP cases.

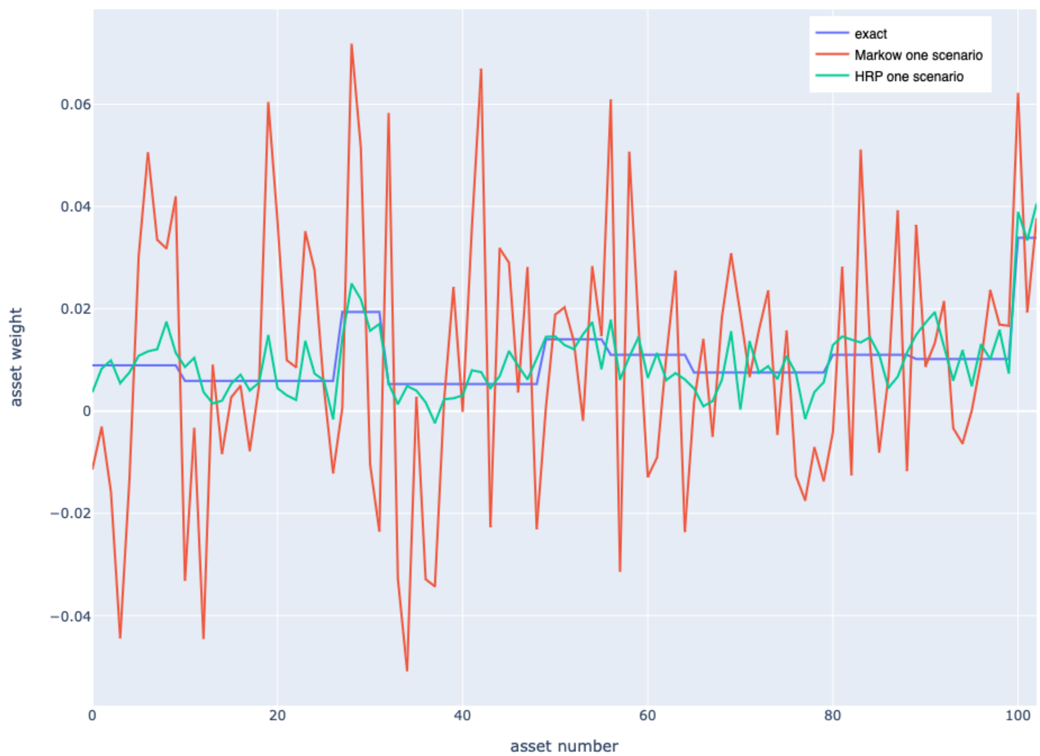

Figure 2: Allocation weights.

For this single scenario we clearly observe a large noise of the Markowitz optimization and a much lower HRP one. To analyse this noise more systematically we will present the following standard deviation per asset, i.e.

$$\sqrt{\frac{\mathbb{E}\left[\Delta w^T \Delta w\right]}{N_A}} \quad \text{and} \quad \sqrt{\frac{\mathbb{E}\left[\Delta u^T \Delta u\right]}{N_A}}$$

for the direct Markowitz and for the HRP, respectively (as stated above, this statistics is estimated for 10,000 Monte Carlo scenarios). This measure can be also thought as an *asset weight confidence interval* for one scenario (e.g. real life one) estimation.

In the table below we will see that the obtained confidence intervals for the Markowitz optimization are (much) larger than the average asset weight $\sim 1\%$. On the other hand, the HRP gives much smaller confidence intervals.

[3]The volatility is set to one for simplicity.

Number of samples	Markowitz Analyt	Markowitz MC	HRP Analyt	HRP MC
250	4.16%	5.42%	1.31%	1.39%
500	2.94%	3.30%	0.93%	0.95%
750	2.40%	2.59%	0.76%	0.77%
1000	2.08%	2.20%	0.66%	0.66%

Table 2: Confidence intervals for asset weights for a zero off-block correlation.

The next table contains a full range of theoretical and Monte Carlo expectations:

$$\mathbb{E}\left[\Delta w^T \Delta w\right] \quad \text{and} \quad \mathbb{E}\left[\Delta u^T \Delta u\right]$$

for the direct Markowitz and for the HRP, respectively.

Off-block corr	Number of samples	Markowitz Analyt	Markowitz MC	HRP Analyt	HRP MC
0	250	0.1785	0.3028	0.0177	0.0198
0.1	250	0.3759	0.6380	0.0177	0.0218
0.2	250	0.5733	0.9740	0.0177	0.0246
0.3	250	0.7710	1.3110	0.0177	0.0287
0.4	250	0.9692	1.6490	0.0177	0.0349
0.5	250	1.1683	1.9890	0.0177	0.0455
0.6	250	1.3695	2.3332	0.0177	0.0671
0	500	0.0893	0.1122	0.0089	0.0093
0.1	500	0.1879	0.2363	0.0089	0.0099
0.2	500	0.2867	0.3606	0.0089	0.0108
0.3	500	0.3855	0.4852	0.0089	0.0120
0.4	500	0.4846	0.6103	0.0089	0.0138
0.5	500	0.5841	0.7362	0.0089	0.0170
0.6	500	0.6848	0.8638	0.0089	0.0238
0	750	0.0595	0.0689	0.0059	0.0061
0.1	750	0.1253	0.1451	0.0059	0.0065
0.2	750	0.1911	0.2214	0.0059	0.0069
0.3	750	0.2570	0.2980	0.0059	0.0075
0.4	750	0.3231	0.3748	0.0059	0.0085
0.5	750	0.3894	0.4520	0.0059	0.0102
0.6	750	0.4565	0.5302	0.0059	0.0139
0	1000	0.0446	0.0497	0.0044	0.0045
0.1	1000	0.0940	0.1047	0.0044	0.0048
0.2	1000	0.1433	0.1597	0.0044	0.0050
0.3	1000	0.1928	0.2149	0.0044	0.0055
0.4	1000	0.2423	0.2702	0.0044	0.0061
0.5	1000	0.2921	0.3259	0.0044	0.0072
0.6	1000	0.3424	0.3822	0.0044	0.0097

Table 3: Analytical and estimated noise for the Markowitz and HRP optimizations.

For better visualization we also plot a *normalize* noise

$$N_T\, \mathbb{E}\left[\Delta w^T \Delta w\right] \quad \text{and} \quad N_T\, \mathbb{E}\left[\Delta u^T \Delta u\right]$$

for the direct Markowitz and for the HRP, respectively.

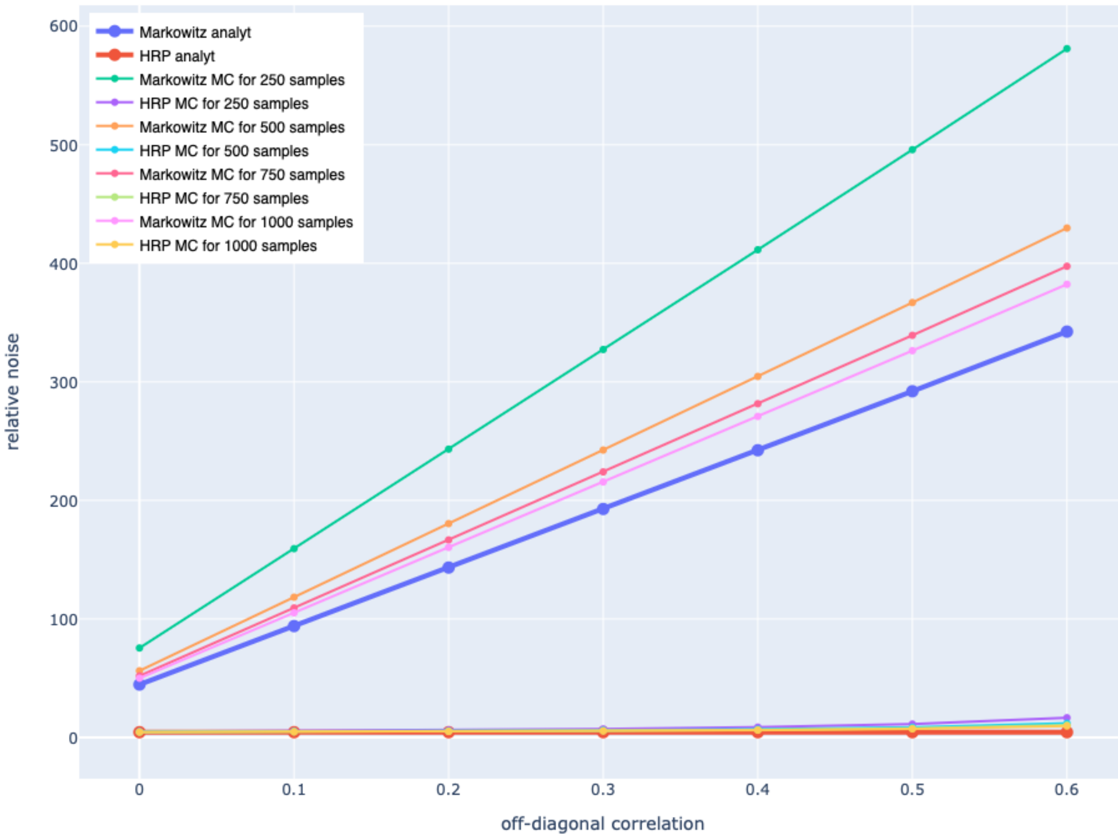

Figure 3: Analytical and estimated relative noise for the Markowitz and HRP optimizations.

We observe a gap between Monte Carlo noise calculation and the analytics for the Markowitz optimization. Its origin is due to non-linear 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.

Similarly, a gap between the HRP Monte Carlo noise calculation and the analytics is due to the non-linearity *but also* the fact that the analytics ignores the off-block elements.

Importantly, we see that *the impact of the HRP to the noise reduction is very significant*: 10 times for a pure block structure and 30 for more significant off-block correlations!

To finalize this subsection we address another noise measure introduced in [5]: a *variance error*

$$\mathbb{E}\left[\Delta w^T \bar{V}\, \Delta w\right] \quad \text{and} \quad \mathbb{E}\left[\Delta u^T \bar{V} \Delta u\right]$$

for Markowitz and HRP methods, respectively. As explained in the previous section, this expression participates in the optimal portfolio expected variance with different signs for the IS and OOS setups, see (27) and (32). The expectations in hand are calculated analytically and numerically. For compactness, we provide errors corresponding to zero off-block correlation.

Number of samples	Markowitz Analyt	Markowitz MC	HRP Analyt	HRP MC
250	0.033	0.056	0.006	0.006
500	0.017	0.021	0.003	0.003
750	0.011	0.013	0.002	0.002
1000	0.008	0.009	0.001	0.001

Table 4: Variance error for a zero off-block correlation.

As in the previous table, we observe here a good fit between the analytics and the MC, as well as a significant noise reduction – as large as 5-10 times – of the HRP w.r.t. the Markowitz optimization.

Our next set of experiments will deal with the risk of the optimal portfolio.

Optimal portfolio variance. We have calculated different statistical characteristics of both IS and OOS portfolios for our two optimizations, Markowitz and HRP. Namely:

- The expected portfolio variance
  - Analytics
    - → Zero order value $\bar{\sigma}^2$ (common for all IS/OOS and Markowitz/HRP) It corresponding to a limit of the large number of samples $N_T \to \infty$
    - → Its first order adjusted values (30-32) The obtained variances do not coincide any more but form the inequality (35)
  - Monte Carlo We calculate the numerical expectation over 10,000 simulation scenarios
- The standard deviation of the portfolio variance
  - Analytics leading order value $\sqrt{2/N_T}\bar{\sigma}^2$ It is common for all IS/OOS and Markowitz/HRP due to (36-37)
  - Monte Carlo We calculate the numerical standard deviation over 10,000 simulation scenarios

Let us start with the **expected portfolio variance** and summarize the results in the table below.

Off block corr	Number of samples	Analyt 0 order	In-sample Markowitz Analyt	In-sample Markowitz MC	In-sample HRP Analyt	In-sample HRP MC	Out-of-sample Markowitz Analyt	Out-of-sample Markowitz MC	Out-of-sample HRP Analyt	Out-of-sample HRP MC
0	250	0.081	0.048	0.048	0.075	0.075	0.115	0.137	0.087	0.088
0.1	250	0.171	0.101	0.101	0.159	0.156	0.241	0.289	0.184	0.181
0.2	250	0.261	0.155	0.155	0.242	0.236	0.368	0.440	0.280	0.274
0.3	250	0.351	0.208	0.208	0.325	0.316	0.494	0.592	0.377	0.367
0.4	250	0.441	0.261	0.261	0.409	0.395	0.621	0.744	0.473	0.461
0.5	250	0.531	0.314	0.314	0.492	0.474	0.747	0.895	0.570	0.556
0.6	250	0.620	0.367	0.367	0.575	0.551	0.873	1.046	0.665	0.652
0	500	0.081	0.065	0.065	0.078	0.078	0.098	0.102	0.084	0.084
0.1	500	0.171	0.136	0.136	0.165	0.163	0.206	0.215	0.178	0.176
0.2	500	0.261	0.208	0.208	0.252	0.248	0.315	0.328	0.271	0.267
0.3	500	0.351	0.280	0.280	0.338	0.333	0.423	0.440	0.364	0.359
0.4	500	0.441	0.351	0.351	0.425	0.418	0.531	0.553	0.457	0.450
0.5	500	0.531	0.422	0.423	0.511	0.503	0.639	0.665	0.550	0.542
0.6	500	0.620	0.494	0.494	0.597	0.586	0.747	0.777	0.643	0.634
0	750	0.081	0.070	0.070	0.079	0.079	0.093	0.094	0.083	0.083
0.1	750	0.171	0.148	0.148	0.167	0.166	0.195	0.198	0.176	0.174
0.2	750	0.261	0.226	0.226	0.255	0.253	0.297	0.302	0.268	0.265
0.3	750	0.351	0.303	0.303	0.343	0.339	0.399	0.406	0.360	0.356
0.4	750	0.441	0.381	0.381	0.430	0.426	0.501	0.510	0.452	0.447
0.5	750	0.531	0.459	0.459	0.518	0.512	0.603	0.614	0.544	0.538
0.6	750	0.620	0.536	0.536	0.605	0.598	0.704	0.717	0.635	0.629
0	1000	0.081	0.073	0.073	0.080	0.080	0.090	0.091	0.083	0.083
0.1	1000	0.171	0.154	0.154	0.168	0.167	0.189	0.191	0.175	0.174
0.2	1000	0.261	0.235	0.235	0.257	0.255	0.288	0.291	0.266	0.264
0.3	1000	0.351	0.315	0.315	0.345	0.342	0.387	0.391	0.358	0.355
0.4	1000	0.441	0.396	0.396	0.433	0.430	0.486	0.491	0.449	0.446
0.5	1000	0.531	0.477	0.477	0.521	0.517	0.585	0.591	0.540	0.536
0.6	1000	0.620	0.557	0.557	0.609	0.603	0.683	0.690	0.631	0.627

Table 5: Portfolio variance expectation.

We observe the following:

- The expected variances do obey the inequality (35) for both Analytical and MC answers. Namely, the smallest value has the IS Markowitz portfolio followed by the IS HRP. Both are less than the theoretical variance. On the other side of the theoretical variance there are the HRP OOS and the Markowitz OOS.
- The HRP values are much closer to the theoretical variance than the Markowitz ones.
- The OOS portfolio variance reduction of the HRP w.r.t. the Markowitz starts with 50% for 250 timesteps (one year interval) and end at 10% for 1000 timesteps.
- The analytics quality is excellent due to its higher order in the number of samples.
- Off-diagonal correlations which pulls us out of our assumptions do not break the picture: the HRP is much more efficient than the Markowitz method.

Below, for a better visualization of the above effects, we present a plot of the expected variance as function of timesteps for all the portfolios for a zero off-block correlation as well as the expected variance as function of different off-block correlations for 500 timesteps.

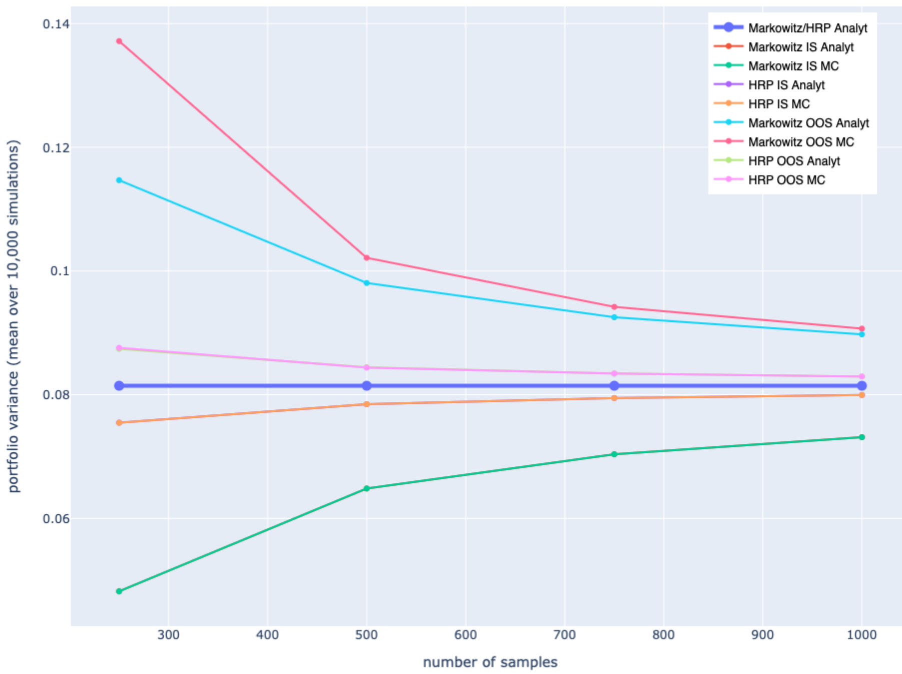

Figure 4: Portfolio variance expectations for zero off-block correlation.

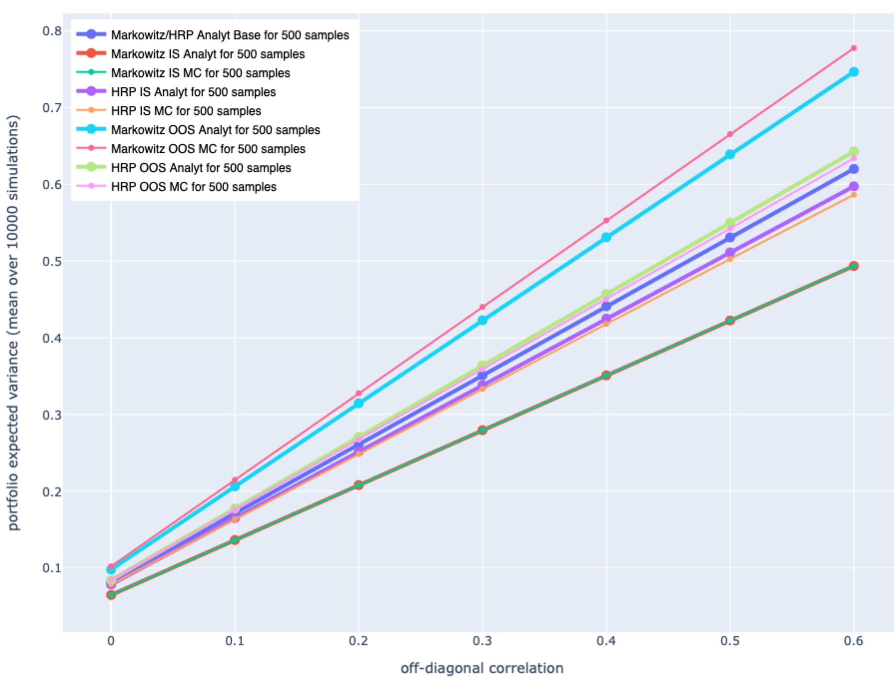

Figure 5: Portfolio variance expectations for 500 samples.

Now we pass to the **standard deviation of the portfolio variance** which we calculate analytically in the leading order and numerically over 10,000 simulations.

		Analyt	Monte Carlo			
Off-block	Number	lead-order	In-sample		Out-of-sample	
corr	of samples		Markowitz	HRP	Markowitz	HRP
0	250	0.007	0.006	0.007	0.016	0.008
0.1	250	0.015	0.012	0.014	0.033	0.016
0.2	250	0.023	0.018	0.022	0.051	0.025
0.3	250	0.031	0.024	0.029	0.069	0.034
0.4	250	0.039	0.030	0.036	0.086	0.042
0.5	250	0.047	0.036	0.044	0.104	0.051
0.6	250	0.055	0.042	0.051	0.122	0.061
0	500	0.005	0.005	0.005	0.007	0.005
0.1	500	0.011	0.010	0.010	0.015	0.011
0.2	500	0.017	0.015	0.016	0.023	0.017
0.3	500	0.022	0.020	0.021	0.031	0.023
0.4	500	0.028	0.025	0.027	0.039	0.029
0.5	500	0.034	0.030	0.032	0.047	0.035
0.6	500	0.039	0.035	0.038	0.055	0.041
0	750	0.004	0.004	0.004	0.005	0.004
0.1	750	0.009	0.008	0.009	0.011	0.009
0.2	750	0.013	0.012	0.013	0.017	0.014
0.3	750	0.018	0.017	0.017	0.023	0.019
0.4	750	0.023	0.021	0.022	0.028	0.023
0.5	750	0.027	0.025	0.026	0.034	0.028
0.6	750	0.032	0.029	0.031	0.040	0.033
0	1000	0.004	0.003	0.004	0.004	0.004
0.1	1000	0.008	0.007	0.007	0.009	0.008
0.2	1000	0.012	0.011	0.011	0.014	0.012
0.3	1000	0.016	0.015	0.015	0.018	0.016
0.4	1000	0.020	0.019	0.019	0.023	0.020
0.5	1000	0.024	0.022	0.023	0.028	0.024
0.6	1000	0.028	0.026	0.027	0.032	0.028

Table 6: Portfolio variance standard deviation.

As in the case of the expected variance, we observe that the portfolio variance noise (standard deviation) satisfies the similar inequality (38) and that the OOS noise of the Markowitz can be substantially lower than the Markowitz one, esp. for low 200 or 500 samples.

We also see that the difference between the expected Markowitz portfolio risk and the HRP one is significantly larger than their standard deviations. This confirms our theoretical conclusion than the difference between the Markowitz portfolio risk and the HRP one for **a single scenario** is mostly due to a difference between their expected values rather than to a standard deviation of the risks.

Below, for a better visualization of the above effects, we present a plot of the standard deviation of the variance as function of samples for all the portfolios for a zero off-block correlation as well as the standard deviation of the variance as function of different off-block correlations for 500 smaples.

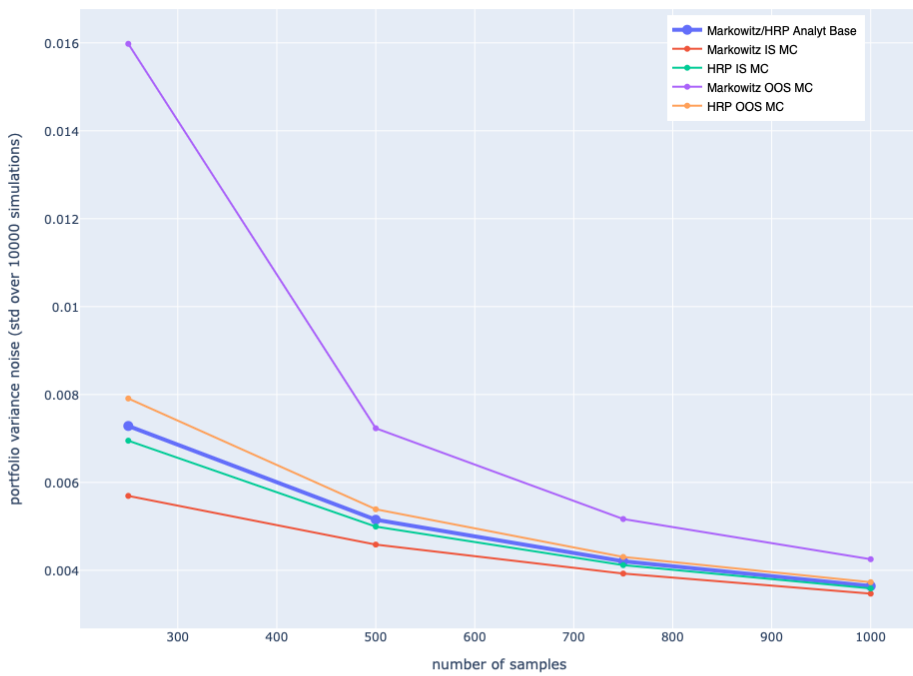

Figure 6: Portfolio variance standard deviation for zero off-block correlation.

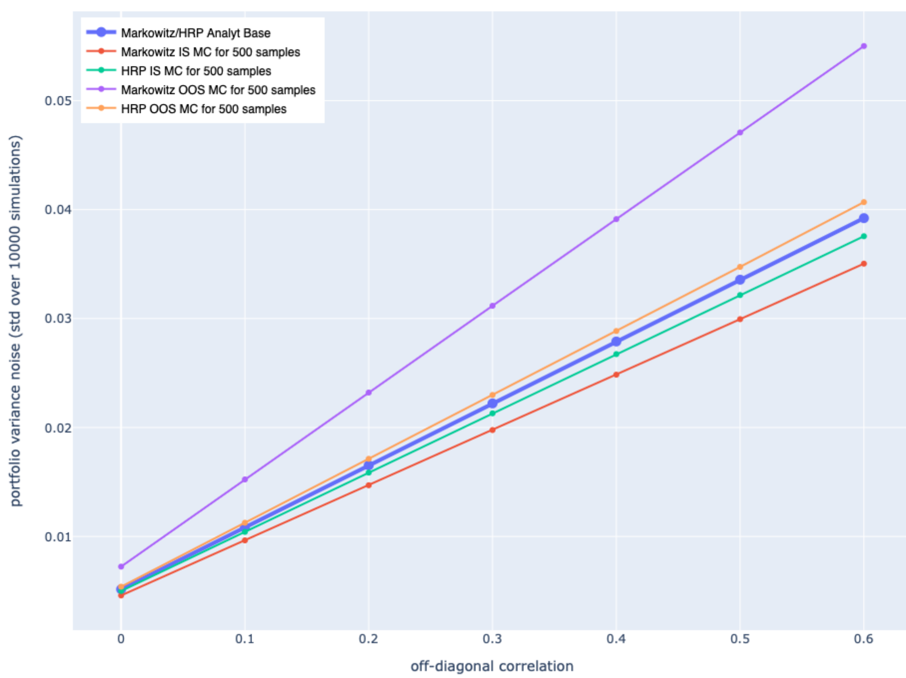

Figure 7: Portfolio variance standard deviation for 500 time-steps.

6 Conclusions

We have derived analytical formulas estimating the noise of portfolio optimization weights for both Markowitz optimization and the HRP approach. Their comparison shows that the HRP is less noisy than the Markowitz and much more robust.

Another important part of the paper was devoted to a detailed analysis of the optimal portfolio variance. For practical applications only the *out-of-sample* setup has value. In this way, we derived the portfolio variance analytical formulas and theoretically demonstrate the superiority of the HRP w.r.t to the Markowitz optimization. The analytics were calculated in a higher order of the number of timesteps which provided an excellent approximation quality. We confirmed the theoretical results using multiple numerical experiments based on Monte Carlo simulations.

Apart from the theoretical evidence of the HRP superiority w.r.t to the Markowitz we have addressed direct practical outcomes of our analytics. The first one was a fast estimation of the *confidence level* of the optimization weights calculated for a single (real-life) scenario. Indeed, given the number of timesteps in the covariance matrix estimation we are able to validate the result out of the noise. The second practical usefulness of the analytics was an HRP portfolio construction criterion which selects assets and clusters minimizing the analytical portfolio variance.

We are grateful to our ADIA colleagues esp. to Adil Reghai for stimulation discussions.

References

[1] Bruno de Finetti (1940) "Il problema dei "Pieni"", *Giornale dell' Istituto Italiano degli Attuari* 11, 1–88; translation "Barone, L. (2006) "The problem of full-risk insurances. Chapter I. The risk within a single accounting period", *Journal of Investment Management* 4(3), 19–43

[2] Vladimir Marchenko and Leonid Pastur (1967) "Distribution of eigenvalues for some sets of random matrices", *Matematicheskii Sbornik,* 114(4), 507–536.

[3] Harry Markowitz (1952) "Portfolio Selection", *The Journal of Finance*, 7, 77-91.

[4] Marcos López de Prado (2016) "Building Diversified Portfolios that Outperform Out of Sample", *The Journal of Portfolio Management*, 42 (4) 59-69

[5] Marcos López de Prado (2016) "A Robust Estimator of the Efficient Frontier", SSRN https://ssrn.com/abstract=3469961

[6] Marcos López de Prado (2020), "Machine Learning for Asset Managers", Cambridge Elements in Quantitative Finance

[7] Flavio Pressacco and Paolo Serafini (2007) "The origins of the mean-variance approach in finance: revisiting de Finetti 65 years later", *Decisions in Economics and Finance* 30, 19–49

A Expected noise for the Markowitz weights.

As shown in the main body of the paper (8), the noise of the weights is

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

Below we will calculate the covariance $\mathbb{E}\left[\Delta w\, \Delta w^T\right]$ which depends on the following expectation

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

For this we will use a general formula for an arbitrary matrix M

$$\mathbb{E}\left[\Delta V\, M\, \Delta V\right] = \frac{1}{N_T}\left(\mathbb{E}\left[X\, X^T\, (X^T\, M\, X)\right] - \bar{V}\, M\, \bar{V}\right)$$

which can be proven as follows

$$\begin{aligned}
\sum_{k,n} \mathbb{E}\left[\Delta V_{ik}\, M_{kn}\, \Delta V_{nj}\right] &= \frac{1}{N_T^2} \sum_{k,n,p,p'} \mathbb{E}\left[\left(X_{ip}\, X_{kp} - \mathbb{E}\left[X_i\, X_k\right]\right) M_{kn} \left(X_{np'}\, X_{jp'} - \mathbb{E}\left[X_n\, X_j\right]\right)\right] \\
&= \frac{1}{N_T^2} \sum_{k,n,p} \mathbb{E}\left[\left(X_{ip}\, X_{kp} - \mathbb{E}\left[X_i\, X_k\right]\right) M_{kn} \left(X_{np}\, X_{jp} - \mathbb{E}\left[X_n\, X_j\right]\right)\right] \\
&= \frac{1}{N_T} \sum_{k,n} \mathbb{E}\left[\left(X_i\, X_k - \mathbb{E}\left[X_i\, X_k\right]\right) M_{kn} \left(X_n\, X_j - \mathbb{E}\left[X_n\, X_j\right]\right)\right] \\
&= \frac{1}{N_T} \sum_{k,n} \mathbb{E}\left[X_i\, X_k\, M_{kn}\, X_n\, X_j\right] - \frac{1}{N_T} \sum_{k,n} \mathbb{E}\left[X_i\, X_k\right]\, M_{kn}\, \mathbb{E}\left[X_n\, X_j\right] \\
&= \frac{1}{N_T} \left(\mathbb{E}\left[X_i\, X_j\, (X^T\, M\, X)\right] - \left(\bar{V}\, M\, \bar{V}\right)_{ij}\right)
\end{aligned}$$

For our concrete case of $M = \bar{w}\, \bar{w}^T$ we obtain

$$\mathbb{E}\left[\Delta V\, \bar{w}\, \bar{w}^T\, \Delta V\right] = \frac{1}{N_T} \left(\mathbb{E}\left[X\, X^T\, (\bar{w}^T\, X)^2\right] - \frac{1}{N_T}\, \bar{V}\, \bar{w}\, \bar{w}^T\, \bar{V}\right)$$

which gives the covariance of the weights noise

$$\begin{aligned}
\mathbb{E}\left[\Delta w\, \Delta w^T\right] &\approx \frac{1}{N_T}(1 - \bar{w}\, a^T) \left(\mathbb{E}\left[\bar{V}^{-1}\, X\, X^T\, \bar{V}^{-1}\, (\bar{w}^T\, X)^2\right] - \bar{w}\, \bar{w}^T\right) (1 - a\, \bar{w}^T) \\
&= \frac{1}{N_T}(1 - \bar{w}\, a^T) \mathbb{E}\left[\bar{V}^{-1}\, X\, X^T\, \bar{V}^{-1}\, (\bar{w}^T\, X)^2\right] (1 - a\, \bar{w}^T) \qquad (39)
\end{aligned}$$

where the second equality is based on the weights constraint $a^T\, \bar{w} = 1$.

The noise covariance depends on *4-points averages* $\mathbb{E}\left[X_n\, X_m\, X_i\, X_j\right]$. Let us diagonalize the covariance matrix

$$\bar{V} = A\, A^T$$

s.t. each return

$$X_n = \sum_{n'} A_{nn'}\, Z_{n'}$$

for normalized Z's, $I = \mathbb{E}\left[Z\, Z^T\right]$. Their 4-points averages are simply

$$\mathbb{E}\left[Z_n\, Z_m\, Z_i\, Z_j\right] = \delta_{nm}\, \delta_{ij} + \delta_{ni}\, \delta_{mj} + \delta_{nj}\, \delta_{mi} + \left(\mathbb{E}\left[Z_n^4\right] - 3\right) \delta_{nm}\, \delta_{mi}\, \delta_{ij}$$

Coming back to our correlated X's 4-points we have

$$\mathbb{E}\left[X_n\, X_m\, X_i\, X_j\right] = \bar{V}_{nm}\, \bar{V}_{ij} + \bar{V}_{ni}\, \bar{V}_{mj} + \bar{V}_{nj}\, \bar{V}_{mi} + \left(\mathbb{E}\left[Z_n^4\right] - 3\right) \sum_{n'} A_{nn'}\, A_{mn'}\, A_{in'}\, A_{jn'} \qquad (40)$$

where the last (smaller) term contains much less summations than the main first three ones. Now, let us come to our main average (39) which can be expressed as

$$\begin{aligned}
\mathbb{E}\left[\bar{V}^{-1}\, X\, X^T\, \bar{V}^{-1}\, (\bar{w}^T\, X)^2\right]_{nm} &= \sum_{'} \bar{V}^{-1}_{nm'}\, \bar{V}^{-1}_{k'm}\, \bar{w}_{j'}\, \bar{w}_{q'}\, \mathbb{E}\left[X_{m'}\, X_{k'}\, X_{j'}\, X_{q'}\right] \\
&= \bar{V}^{-1}_{nm}\, (\bar{w}^T\, \bar{V}\, \bar{w}) + 2\, \bar{w}_n\, \bar{w}_m \\
&+ \left(\mathbb{E}\left[Z_n^4\right] - 3\right) \sum_{n'} (\bar{V}^{-1}\, A)_{nn'}\, (\bar{V}^{-1}\, A)_{mn'}\, (a^T\, \bar{V}^{-1}\, A)^2_{n'}
\end{aligned}$$

In what follows we will assume the Gaussian nature of the returns which permit us to ignore the last term: it is exactly zero for Gaussian Z's. Given the above can easily derive a compact expression for the noise matrix [4]

$$\mathbb{E}\left[\Delta w\, \Delta w^T\right] \approx \frac{1}{N_T}(1 - \bar{w}\, a^T) \left(\frac{\bar{V}^{-1}}{\Omega} + 2\bar{w}\, \bar{w}^T\right) (1 - a\, \bar{w}^T) = \frac{1}{N_T} \left(\frac{\bar{V}^{-1}}{\Omega} - w\, w^T\right) \qquad (41)$$

where Ω is the following quadratic form

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

[4] We have used again $a^T\, \bar{w} = 1$.

Similarly, we can derive a useful general formula

$$\mathbb{E}\left[\Delta V\, M\, \Delta V\right] = \frac{1}{N_T}\left(\bar{V}\,\mathrm{Tr}(\bar{V}\,M) + \bar{V}\,M^T\,\bar{V}\right)$$

valid for the Gaussian case which we will use in the main body of the paper.

B The total noise for the HRP.

In this appendix we derive the expected noise formula for the HRP. Let us start with its off-block part.

Off-block noise for the HRP. Denote diagonal block quantities with a subscript B : K_B is the diagonal part of the clusters' covariance matrix

$$\left(K_B^{-1}\right)_{hq} = \delta_{hq}\,\Omega_h$$

The corresponding cluster weights read

$$(\xi_B)_h = \frac{\left(K_B^{-1}\,\iota\right)_h}{\iota^T\,K_B^{-1}\,\iota} = \frac{\Omega_h}{\Omega} \tag{42}$$

where $\Omega = \sum_h \Omega_h$. The off-diagonal noise in the ξ-weights, which we denote as

$$\delta\xi \equiv \xi - \xi_B,$$

corresponds to "Monte Carlo movements" of the clusters' covariance matrix out of its block diagonal

$$\delta K \equiv K - K_B$$

Applying the formula (8) to the optimal cluster weights ξ_B gives the noise

$$\delta\xi \approx -\left(I - \xi_B\,\iota^T\right) K_B{}^{-1}\delta K\,\xi_B$$

which comes from the Monte Carlo noise of the covariance matrix

$$\delta K_{hq} = 1_{h\neq q} w^{(h)^T}\,\delta V^{(h,q)} w^{(q)} \tag{43}$$

where

$$\left(\delta V^{(h,q)}\right)_{nm} = \frac{1}{N_T}\sum_{p=1}^{N_T} Y_{n,p}^{(h)}\,Y_{m,p}^{(q)}$$

Thus, the off-diagonal noise of the final weights $u^{(h)} = \xi_h\,w^{(h)}$ is only due to that inside ξ's

$$\begin{aligned}
\delta u^{(h)} &= w^{(h)}\,\delta\xi_h = -w^{(h)}\sum_{rq}\left(\delta_{hr} - \frac{\Omega_h}{\Omega}\right)\Omega_r\,\delta K_{rq}\,\frac{\Omega_q}{\Omega} && (44)\\
&= -w^{(h)}\sum_{r\neq q}\left(\delta_{hr} - \frac{\Omega_h}{\Omega}\right)\frac{\Omega_r\,\Omega_q}{\Omega}\;w^{(r)^T}\,\delta V^{(r,q)} w^{(q)} && (45)
\end{aligned}$$

The cluster optimization noise[5] depends on a quadratic form expectation of the matrix δV

$$\begin{aligned}
\mathcal{N}_C' &\equiv \sum_h \mathbb{E}\left[\delta u^{(h)^T}\,\delta u^{(h)}\right]\\
&\approx \sum_h w^{(h)^T}\,w^{(h)}\sum_{r\neq q}\sum_{r'\neq q'}\left(\delta_{hr} - \frac{\Omega_h}{\Omega}\right)\left(\delta_{hr'} - \frac{\Omega_h}{\Omega}\right)\frac{\Omega_r\,\Omega_q}{\Omega}\;\frac{\Omega_{r'}\,\Omega_{q'}}{\Omega}\\
&\times\; w^{(r)^T}\,\mathbb{E}\left[\delta V^{(r,q)}\,w^{(q)}\,w^{(r')^T}\,\delta V^{(r',q')}\right]\,w^{(q')}
\end{aligned}$$

[5] The prime symbol in the notations corresponds to the off-block part.

The underlying expectation can be rewritten as

$$
\begin{aligned}
& \mathbb{E}\left[w^{(r)^T}\,\delta V^{(r,q)}\,w^{(q)}\,w^{(r')^T}\,\delta V^{(r',q')} w^{(q')}\right]\delta_{r\neq q}\,\delta_{r'\neq q'} \\
= \quad & \frac{1}{N_T}\mathbb{E}\left[w^{(r)^T}\,Y^{(r)}\,Y^{(q)^T}\,w^{(q)}\,w^{(r')^T}\,Y^{(r')}\,Y^{(q')^T}\,w^{(q')}\right]\delta_{r\neq q}\,\delta_{r'\neq q'} \\
= \quad & \frac{1}{N_T}\mathbb{E}\left[w^{(r)^T}\,Y^{(r)}\,Y^{(r')^T}\,w^{(r')}\right]\mathbb{E}\left[w^{(q)^T}\,Y^{(q)}\,Y^{(q')^T}\,w^{(q')}\right]\delta_{r=r'}\,\delta_{q=q'}\,\delta_{r\neq q}\,\delta_{r'\neq q'} \\
+ \quad & \frac{1}{N_T}\mathbb{E}\left[w^{(r)^T}\,Y^{(r)}\,Y^{(q')^T}\,w^{(q')}\right]\mathbb{E}\left[w^{(q)^T}\,Y^{(q)}\,Y^{(r')^T}\,w^{(r')}\right]\delta_{r=q'}\,\delta_{q=r'}\,\delta_{r\neq q}\,\delta_{r'\neq q'} \\
= \quad & \frac{1}{N_T}\mathbb{E}\left[w^{(r)^T}\,Y^{(r)}\,Y^{(r)^T}\,w^{(r)}\right]\mathbb{E}\left[w^{(q)^T}\,Y^{(q)}\,Y^{(q)^T}\,w^{(q)}\right]\left(\delta_{r=r'}\,\delta_{q=q'}+\delta_{r=q'}\,\delta_{q=r'}\right)\delta_{r\neq q}\,\delta_{r'\neq q'} \\
\simeq \quad & \frac{1}{N_T}\Omega_r^{-1}\,\Omega_q^{-1}\left(\delta_{r=r'}\,\delta_{q=q'}+\delta_{r=q'}\,\delta_{q=r'}\right)\delta_{r\neq q}\,\delta_{r'\neq q'}
\end{aligned}
$$

where we have used

$$
\mathbb{E}\left[w^{(r)^T}\,Y^{(r)}\,Y^{(r)^T}\,w^{(r)}\right]=\mathbb{E}\left[w^{(r)^T}\,\bar{V}^{(r)}\,w^{(r)}\right]\simeq\Omega_r^{-1}
$$

and approximated $\bar{V}^{(r)}$ with $V^{(r)}+O(N_T^{-1})$. Inserting it into the formula for N'_C we obtain

$$
\begin{aligned}
\mathcal{N}'_C \quad \approx \quad & \frac{1}{N_T}\sum_h w^{(h)^T}\,w^{(h)}\sum_{r\neq q}\sum_{r'\neq q'}\left(\delta_{hr}-\frac{\Omega_h}{\Omega}\right)\left(\delta_{hr'}-\frac{\Omega_h}{\Omega}\right)\frac{\Omega_r\,\Omega_q}{\Omega}\;\frac{\Omega_{r'}\,\Omega_{q'}}{\Omega} \\
\times \quad & \Omega_r^{-1}\,\Omega_q^{-1}\left(\delta_{r=r'}\,\delta_{q=q'}+\delta_{r=q'}\,\delta_{q=r'}\right) \\
= \quad & \frac{1}{N_T}\sum_h w^{(h)^T}\,w^{(h)}\sum_{r\neq q}\left(\delta_{hr}-\frac{\Omega_h}{\Omega}\right)^2\frac{\Omega_r\,\Omega_q}{\Omega}\;\frac{\Omega_r\,\Omega_q}{\Omega}\Omega_r^{-1}\,\Omega_q^{-1} \\
+ \quad & \frac{1}{N_T}\sum_h w^{(h)^T}\,w^{(h)}\sum_{r\neq q}\left(\delta_{hr}-\frac{\Omega_h}{\Omega}\right)\left(\delta_{hq}-\frac{\Omega_h}{\Omega}\right)\frac{\Omega_r\,\Omega_q}{\Omega}\;\frac{\Omega_q\,\Omega_r}{\Omega}\Omega_r^{-1}\,\Omega_q^{-1} \\
= \quad & \frac{1}{N_T}\sum_h w^{(h)^T}\,w^{(h)}\sum_{r\neq q}\left(\delta_{hr}-\frac{\Omega_h}{\Omega}\right)\left(\delta_{hq}+\delta_{hr}-2\frac{\Omega_h}{\Omega}\right)\frac{\Omega_r\,\Omega_q}{\Omega^2} \\
= \quad & \frac{1}{N_T}\sum_h w^{(h)^T}\,w^{(h)}\sum_{r\neq q}\left(\delta_{hr}-\frac{\Omega_h}{\Omega}\left(3\delta_{hr}+\delta_{hq}\right)+2\left(\frac{\Omega_h}{\Omega}\right)^2\right)\frac{\Omega_r\,\Omega_q}{\Omega^2} \\
= \quad & \frac{1}{N_T}\sum_h w^{(h)^T}\,w^{(h)}\sum_{r,q}\left(\delta_{hr}-\frac{\Omega_h}{\Omega}\left(3\delta_{hr}+\delta_{hq}\right)+2\left(\frac{\Omega_h}{\Omega}\right)^2\right)\frac{\Omega_r\,\Omega_q}{\Omega^2} \\
- \quad & \frac{1}{N_T}\sum_h w^{(h)^T}\,w^{(h)}\sum_{r}\left(\delta_{hr}-4\frac{\Omega_h}{\Omega}\,\delta_{hr}+2\left(\frac{\Omega_h}{\Omega}\right)^2\right)\frac{\Omega_r^2}{\Omega^2} \\
= \quad & \frac{1}{N_T}\sum_h w^{(h)^T}\,w^{(h)}\left(\frac{\Omega_h}{\Omega}-4\left(\frac{\Omega_h}{\Omega}\right)^2+2\left(\frac{\Omega_h}{\Omega}\right)^2\right) \\
- \quad & \frac{1}{N_T}\sum_h w^{(h)^T}\,w^{(h)}\left(\frac{\Omega_h^2}{\Omega^2}-4\frac{\Omega_h^3}{\Omega^3}+2\left(\frac{\Omega_h}{\Omega}\right)^2\sum_r\frac{\Omega_r^2}{\Omega^2}\right) \\
= \quad & \frac{1}{N_T}\sum_h w^{(h)^T}\,w^{(h)}\left(\frac{\Omega_h}{\Omega}-3\left(\frac{\Omega_h}{\Omega}\right)^2+4\frac{\Omega_h^3}{\Omega^3}-2\left(\frac{\Omega_h}{\Omega}\right)^2\sum_r\frac{\Omega_r^2}{\Omega^2}\right)
\end{aligned}
$$

Finally, using $\Omega_h^2\,w^{(h)^T}\,w^{(h)}=a^{(h)^T}\,V^{(h)^{-2}}\,a^{(h)}$ we obtain the off-block clustered optimization noise

$$
\mathcal{N}'_C \quad \simeq \quad \frac{1}{N_T}\frac{1}{\Omega}\sum_h\frac{a^{(h)^T}\,V^{(h)^{-2}}\,a^{(h)}}{\Omega_h}\left(1-3\frac{\Omega_h}{\Omega}+4\frac{\Omega_h^2}{\Omega^2}-2\frac{\Omega_h}{\Omega}\sum_r\frac{\Omega_r^2}{\Omega^2}\right) \tag{46}
$$

Next, if order to prove that the HRP is less noisy than the classical Markowitz, we pass to the off-block noise calculations for the latter.

Off-block noise for the Markowitz optimization. To get the Markowitz *off-block* noise we proceed as follows: instead of expanding around the exact covariance matrix $\bar{V}$ (as in Section 2) we will expand around the *block* matrix V_B. To do it we simply replace everywhere $\bar{V}$ by V_B and $\bar{w}$ by w_B defined by (3)

$$w_B = \frac{V_B^{-1}\, a}{a^T\, V_B^{-1}\, a} \tag{47}$$

It is easy to see that these "block" weights coincide with the cluster-optimization weights with the block ξ's (42)

$$\left(\xi_{B\,1}\, w_1^{(1)} \cdots \xi_{B\,1} w_{N_1}^{(1)} \,|\, \cdots \,|\, \xi_{B\,H} w_1^{(H)}, \cdots, \xi_{B\,H} w_{N_H}^{(H)}\right)$$

s.t.

$$w_B^{(h)} = w^{(h)}\, \frac{\Omega_h}{\Omega}$$

The noise of the Markowitz weights over w_B can be calculated as in (8)

$$w - w_B \equiv \delta w_B \approx -(I - w_B\, a^T)\, V_B^{-1}\, \delta V_B\, w_B \tag{48}$$

For a cluster h we have

$$\delta w_B^{(h)} \approx \sum_{r' \neq q'} - \left(\delta_{hr'} - w^{(h)}\, a^{(r')\,T}\, \frac{\Omega_h}{\Omega}\right)\, {V^{(r')}}^{-1}\, \delta V^{(r',q')}\, w^{(q')}\, \frac{\Omega_{q'}}{\Omega}$$

Then, to calculate the expectation $\mathbb{E}\left[\delta {w_B^{(h)}}^T\, \delta w_B^{(h)}\right]$ we notice that

$$\begin{aligned}
&\mathbb{E}\left[\delta {V^{(r,q)}}^T\, M\, \delta V^{(r',q')}\right]_{nm} \delta_{r\neq q}\, \delta_{r'\neq q'} = \frac{1}{N_T}\mathbb{E}\left[\left(Y^{(r)}\, {Y^{(q)}}^T\right)^T\; M\, Y^{(r')}\, {Y^{(q')}}^T\right]_{nm} \delta_{r\neq q}\, \delta_{r'\neq q'} \\
= \;& \frac{1}{N_T} \sum_{n',m'} \mathbb{E}\left[Y_{n'}^{(r)}\, Y_n^{(q)}\, M_{n',m'}\, Y_{m'}^{(r')}\, Y_m^{(q')}\right] \delta_{r\neq q}\, \delta_{r'\neq q'} \\
= \;& \frac{1}{N_T} \sum_{n',m'} \mathbb{E}\left[Y_{n'}^{(r)}\, Y_{m'}^{(r')}\right]\, M_{n',m'}\, \mathbb{E}\left[Y_n^{(q)}\, Y_m^{(q')}\right] \delta_{r=r'}\, \delta_{q=q'}\, \delta_{r\neq q}\, \delta_{r'\neq q'} \\
+ \;& \frac{1}{N_T} \sum_{n',m'} \mathbb{E}\left[Y_{n'}^{(r)}\, Y_m^{(q')}\right]\, M_{n',m'}\, \mathbb{E}\left[Y_{m'}^{(r')} Y_n^{(q)}\right] \delta_{r=q'}\, \delta_{q=r'}\, \delta_{r\neq q}\, \delta_{r'\neq q'} \\
= \;& \frac{1}{N_T} \sum_{n',m'} \mathbb{E}\left[Y_{n'}^{(r)}\, Y_{m'}^{(r)}\right]\, M_{n',m'}\, \mathbb{E}\left[Y_n^{(q)}\, Y_m^{(q)}\right] \delta_{r=r'}\, \delta_{q=q'}\, \delta_{r\neq q}\, \delta_{r'\neq q'} \\
+ \;& \frac{1}{N_T} \sum_{n',m'} \mathbb{E}\left[Y_{n'}^{(r)}\, Y_m^{(r)}\right]\, M_{n',m'}\, \mathbb{E}\left[Y_{m'}^{(q)} Y_n^{(q)}\right] \delta_{r=q'}\, \delta_{q=r'}\, \delta_{r\neq q}\, \delta_{r'\neq q'} \\
\simeq \;& \frac{1}{N_T} \sum_{n',m'} V_{n',m'}^{(r)}\, M_{n',m'}\, V_{n,m}^{(q)} \delta_{r=r'}\, \delta_{q=q'}\, \delta_{r\neq q}\, \delta_{r'\neq q'} + \frac{1}{N_T} \sum_{n',m'} V_{n'm}^{(r)}\, M_{n',m'}\, V_{m',n}^{(q)} \delta_{r=q'}\, \delta_{q=r'}\, \delta_{r\neq q}\, \delta_{r'\neq q'} \\
= \;& \frac{1}{N_T} \mathrm{tr}\left(M V^{(r)}\right) V_{n,m}^{(q)}\, \delta_{r=r'}\, \delta_{q=q'}\, \delta_{r\neq q}\, \delta_{r'\neq q'} + \frac{1}{N_T}\left(V^{(q)} M^T\, V^{(r)}\right)_{n,m} \delta_{r=q'}\, \delta_{q=r'}\, \delta_{r\neq q}\, \delta_{r'\neq q'}
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathcal{N}'_M \;&=\; \sum_h \mathbb{E}\left[\delta {w_B^{(h)}}^T\, \delta w_B^{(h)}\right] = \sum_h \sum_{r\neq q} \sum_{r'\neq q'} \frac{\Omega_q \Omega_{q'}}{\Omega^2} \\
&\times\; {w^{(q)}}^T \mathbb{E}\left[\delta {V^{(r,q)}}^T\, {V^{(r)}}^{-1}\, \left(\delta_{hr} - a^{(r)} {w^{(h)}}^T\, \frac{\Omega_h}{\Omega}\right) \left(\delta_{hr'} - w^{(h)}\, a^{(r')\,T}\, \frac{\Omega_h}{\Omega}\right)\, {V^{(r')}}^{-1}\, \delta V^{(r',q')}\right] w^{(q')} \\
&=\; \sum_h \sum_{r\neq q} \sum_{r'\neq q'} \frac{\Omega_q \Omega_{q'}}{\Omega^2} {w^{(q)}}^T \mathbb{E}\left[\delta {V^{(r,q)}}^T\, M^{(h,r,r')}\, \delta V^{(r',q')}\right] w^{(q')}
\end{aligned}$$

where we have denoted the matrix M as

$$M^{(h,r,q)} = {V^{(r)}}^{-1}\, \left(\delta_{hr} - a^{(r)} {w^{(h)}}^T\, \frac{\Omega_h}{\Omega}\right) \left(\delta_{hq} - w^{(h)}\, a^{(q)\,T}\, \frac{\Omega_h}{\Omega}\right) {V^{(q)}}^{-1}$$

Next, we have

$$\begin{aligned}
\mathcal{N}'_M &\simeq \frac{1}{N_T}\sum_h\sum_{r\neq q}\sum_{r'\neq q'}\frac{\Omega_q\Omega_{q'}}{\Omega^2}w^{(q)T}\left(\mathrm{tr}\left(M^{(h,r,r')}V^{(r)}\right)V^{(q)}\,\delta_{r=r'}\,\delta_{q=q'}+\left(V^{(q)}M^{(h,r,r')^T}V^{(r)}\right)\delta_{r=q'}\,\delta_{q=r'}\right)w^{(q')}\\
&= \frac{1}{N_T}\sum_h\sum_{r\neq q}\sum_{r'\neq q'}\frac{\Omega_q\Omega_{q'}}{\Omega^2}w^{(q)T}\,\mathrm{tr}\left(M^{(h,r,r')}V^{(r)}\right)V^{(q)}\,\delta_{r=r'}\,\delta_{q=q'}w^{(q')}\\
&+ \frac{1}{N_T}\sum_h\sum_{r\neq q}\sum_{r'\neq q'}\frac{\Omega_q\Omega_{q'}}{\Omega^2}w^{(q)T}\left(V^{(q)}M^{(h,r,r')^T}V^{(r)}\right)\delta_{r=q'}\,\delta_{q=r'}w^{(q')}\\
&= \frac{1}{N_T}\sum_h\sum_{r\neq q}\frac{\Omega_q^2}{\Omega^2}\mathrm{tr}\left(M^{(h,r,r)}V^{(r)}\right)w^{(q)T}V^{(q)}\,w^{(q)}+\frac{1}{N_T}\sum_h\sum_{r\neq q}\frac{\Omega_q\Omega_r}{\Omega^2}w^{(q)T}\left(V^{(q)}M^{(h,r,q)^T}V^{(r)}\right)w^{(r)}\\
&= \frac{1}{N_T}\sum_h\sum_{r\neq q}\frac{\Omega_q^2}{\Omega^2}\mathrm{tr}\left(V^{(r)-1}\left(\delta_{hr}-a^{(r)}w^{(h)T}\frac{\Omega_h}{\Omega}\right)\left(\delta_{hr}-w^{(h)}\,a^{(r)T}\frac{\Omega_h}{\Omega}\right)\right)\frac{1}{\Omega_q}\\
&+ \frac{1}{N_T}\sum_h\sum_{r\neq q}\frac{\Omega_q\Omega_r}{\Omega^2}w^{(r)T}\left(V^{(r)}M^{(h,r,q)}\,V^{(q)}\right)w^{(q)}
\end{aligned}$$

Finally, expanding the product we obtain

$$\begin{aligned}
\mathcal{N}'_M &\simeq \frac{1}{N_T}\sum_h\sum_{r\neq q}\frac{\Omega_q}{\Omega^2}\mathrm{tr}\left(V^{(r)-1}\left(\delta_{hr}-\delta_{hr}a^{(r)}w^{(h)T}\frac{\Omega_h}{\Omega}-\delta_{hr}w^{(h)}\,a^{(r)T}\frac{\Omega_h}{\Omega}+a^{(r)}w^{(h)T}\,w^{(h)}\,a^{(r)T}\frac{\Omega_h^2}{\Omega^2}\right)\right)\\
&+ \frac{1}{N_T}\sum_h\sum_{r\neq q}\frac{\Omega_q\Omega_r}{\Omega^2}w^{(r)T}\left(\delta_{hr}-a^{(r)}w^{(h)T}\frac{\Omega_h}{\Omega}\right)\left(\delta_{hq}-w^{(h)}\,a^{(q)T}\frac{\Omega_h}{\Omega}\right)w^{(q)}\\
&= \frac{1}{N_T}\sum_h\sum_{r\neq q}\frac{\Omega_q}{\Omega^2}\left(\delta_{hr}\left(\mathrm{tr}\left(V^{(r)-1}\right)-2\frac{a^{(h)T}\,V^{(h)-2}\,a^{(h)}}{\Omega}\right)+\frac{a^{(h)T}\,V^{(h)-2}\,a^{(h)}}{\Omega}\frac{\Omega_r}{\Omega}\right)\\
&+ \frac{1}{N_T}\sum_h\sum_{r\neq q}\frac{\Omega_q\Omega_r}{\Omega^2}\left(-\delta_{hr}-\delta_{hq}+\frac{\Omega_h}{\Omega}\right)\frac{a^{(h)T}\,V^{(h)-2}\,a^{(h)}}{\Omega}\frac{1}{\Omega_h}\\
&= \frac{1}{N_T}\sum_h\sum_{r\neq q}\frac{\Omega_q}{\Omega^2}\left(\delta_{hr}\mathrm{tr}\left(V^{(h)-1}\right)+\left(-2\delta_{hr}+\frac{\Omega_r}{\Omega}\right)\frac{a^{(h)T}\,V^{(h)-2}\,a^{(h)}}{\Omega}\right)\\
&+ \frac{1}{N_T}\sum_h\frac{a^{(h)T}\,V^{(h)-2}\,a^{(h)}}{\Omega}\sum_{r\neq q}\frac{\Omega_q\Omega_r}{\Omega^2}\left(-2\delta_{hr}+\frac{\Omega_h}{\Omega}\right)\frac{a^{(h)T}\,V^{(h)-2}\,a^{(h)}}{\Omega}\frac{1}{\Omega_h}\\
&= \frac{1}{N_T}\sum_h\sum_{q\neq h}\frac{\Omega_q}{\Omega^2}\mathrm{tr}\left(V^{(h)-1}\right)+\frac{1}{N_T}\sum_h\frac{a^{(h)T}\,V^{(h)-2}\,a^{(h)}}{\Omega}\sum_{r\neq q}\left(-4\delta_{hr}\frac{\Omega_q}{\Omega^2}+2\frac{\Omega_q\Omega_r}{\Omega^3}\right)\\
&= \frac{1}{N_T}\frac{1}{\Omega}\sum_h\mathrm{tr}\left(V^{(h)-1}\right)\left(1-\frac{\Omega_h}{\Omega}\right)-2\frac{1}{N_T}\frac{1}{\Omega}\sum_h\frac{a^{(h)T}\,V^{(h)-2}\,a^{(h)}}{\Omega}\left(1-2\frac{\Omega_h}{\Omega}+2\sum_q\frac{\Omega_q^2}{\Omega^2}\right)
\end{aligned} \tag{49}$$

The total noise for the HRP optimization. The total noise inside the clustered optimization weights

$$u^{(h)}=\xi_h\,w^{(h)}$$

can be decomposed (in the first order) as follows

$$\Delta u^{(h)}=u^{(h)}-\bar{u}^{(h)}=\left((\xi_B)_h+\delta\xi_h\right)\,w^{(h)}-\bar{u}^{(h)}\simeq\left((\xi_B)_h\,w^{(h)}-\bar{u}^{(h)}\right)+\delta\xi_h\,\bar{w}^{\,(h)}=\Delta\left((\xi_B)_h\,w^{(h)}\right)+\delta\xi_h\,\bar{w}^{\,(h)}$$

where $\Delta\left((\xi_B)_h\,w^{(h)}\right)$ is the block-diagonal noise.

Using the theoretical values of the weights (denoted with the bar symbol)

$$\left(\bar{\xi}_B\right)_h=\frac{\bar{\Omega}_h}{\bar{\Omega}}\quad\text{and}\quad\bar{w}^{(h)}=\frac{\bar{V}^{\,(h)^{-1}}\,a^{(h)}}{a^{(h)T}\,\bar{V}^{\,(h)^{-1}}\,a^{(h)}}$$

permits us to calculate the *full* theoretical clustered optimization weights

$$\bar{u}^{(h)}=\frac{\bar{V}^{\,(h)^{-1}}\,a^{(h)}}{\bar{\Omega}}$$

We immediately see that they coincide with the Markowitz ones (7), i.e.

$$\bar{u} = \bar{w}$$

.

As we have seen above, the off-block and diagonal-block noises are independent, s.t. their expectations can be summed up

$$\mathbb{E}\left[\Delta u^{(h)^T}\,\Delta u^{(h)}\right] = \mathbb{E}\left[\delta\xi_h\,\bar{w}^{\,(h)^T}\,\bar{w}^{\,(h)}\,\delta\xi_h\right] + \mathbb{E}\left[\Delta\left((\xi_B)_h\,w^{(h)}\right)^T\Delta\left((\xi_B)_h\,w^{(h)}\right)\right]$$

Although we can calculate directly the diagonal-block expectation

$$\mathbb{E}\left[\Delta\left((\xi_B)_h\,w^{(h)}\right)^T\Delta\left((\xi_B)_h\,w^{(h)}\right)\right] \tag{50}$$

it will simpler to derive it from the similar block/off-block decomposition for the standard Markowitz optimization. Namely, we decompose the Markowitz noise into two parts

$$\Delta w = w - \bar{w} = w - w_B + w_B - \bar{w} = \delta w_B + \Delta w_B$$

As in the clustered case, these noises are independent and their variances can be treated separately. Indeed, we have already calculated the off-block noise $\mathbb{E}\left[\delta w_B^T\,\delta w_B\right]$ in (49) as well as the total noise (11). Thus, the diagonal-block noise can be obtained by subtraction of (49) from (21)

$$\mathbb{E}\left[\Delta w_B^T\,\Delta w_B\right] \approx \frac{1}{N_T}\frac{1}{\Omega}\left(\sum_h \operatorname{tr}\left(V^{(h)^{-1}}\right)\frac{\Omega_h}{\Omega} + \sum_h \frac{a^{(h)^T}\,V^{(h)^{-2}}\,a^{(h)}}{\Omega}\left(1 - 4\frac{\Omega_h}{\Omega} + 2\sum_q \frac{\Omega_q^2}{\Omega^2}\right)\right)$$

As we mentioned it coincides with the HRP diagonal-block noise (50), such that the final formula for the total HRP noise (19) can be obtained summing up the diagonal-block noise and the off-diagonal one (46).