IMPROVED PORTFOLIO DIVERSIFICATION THROUGH UNSUPERVISED LEARNING

Michael Lewis

This version: August 6, 2019

IMPROVED PORTFOLIO DIVERSIFICATION THROUGH UNSUPERVISED LEARNING

ABSTRACT

This paper introduces the Recursive Clustering Risk Parity (RCRP) approach. The RCRP method builds on the philosophy first introduced in Hierarchical Risk Parity (HRP) in López de Prado [2016], leveraging unsupervised learning to recursively build am optimal portfolio.

HRP introduced the concept of building a diversified portfolio using the inherent structure of the covariance matrix, utilizing graph theory and hierarchical clustering. In doing so, HRP avoided inverting the covariance matrix, a numerically unstable procedure when performed on the notoriously ill-conditioned, if not singular, covariance matrix. However, the procedure relies on sorting the underlying instruments and, in doing so, creating a quasi-diagonalization of the matrix, thereby foregoing the underlying tree structure. RCRP builds on this premise by utilizing this underlying tree structure, leveraging improved clustering techniques and inversion approximations to enhance overall performance.

Keywords: Risk parity, tree graph, cluster, recursive, inverse rank-one update, metric space.

JEL Classification: G0, G1, G2.

AMS Classification: 91G10, 91G60, 91G70, 60E.

PORTFOLIO OPTIMIZATION: A BACKGROUND

This paper proposes a method to obtain two desirable portfolios: in particular, the minimum variance portfolio and the maximum sharpe ratio portfolio.

For completeness, a portfolio is a vector of weights w across N market instruments such that the weights sum to unity, or $1^T w = 1$. Given the covariance matrix Σ, expected excess returns vector μ, and portfolio weights w, the variance of the portfolio is $w^T \Sigma w$ and the expected excess return of the portfolio is $w^T \mu$.

With that terminology in place, we will be discussing the minimum variance portfolio, which satisfies equation (1),

$$argmin_{\,w} w^T \Sigma w$$

$$such\ that\ 1^T w = 1 \quad (1)$$

and the maximum sharpe ratio portfolio, which satisfies equation (2).

$$argmax_{\,w} \frac{w^T \mu}{\sqrt{w^T \Sigma w}}$$

$$such\ that\ 1^T w = 1 \quad (2)$$

The known optimal solution to equation (1) is setting $w = \frac{\Sigma^{-1} 1}{1^T \Sigma^{-1} 1}$, or equivalently $w \propto \Sigma^{-1} 1$, whereas the known optimal solution to equation (2) is setting $w \propto \Sigma^{-1} \mu$. Put more concisely, the solutions set $w \propto \Sigma^{-1} a$, where $a = 1$ for the minimum variance portfolio, and

$a = \mu$ for the max sharpe ratio. Unfortunately, both solutions require inverting the covariance matrix, a numerically unstable task when it's possible, and impossible when the covariance matrix is singular, as is often the case in practice when involving thousands of stocks. Aside from this inherent numerical instability, empirical covariance matrices can themselves be nosy when evaluated on a finite number of financial returns, further complicating matters. If for no other reasons, a method that is robust to noise and does not require ill-conditioned matrix inversion is desirable.

One notable option for the minimum variance portfolio is the inverse variance portfolio. In this case, the weight w_i given to instrument i is set to be inversely proportional to its variance $\sigma_i^2 = \Sigma_{ii}$, or $w_i \propto \frac{1}{\sigma_i^2}$. Observe that this is the optimal solution in the special case where the instruments are independent, or equivalently $\Sigma = \Sigma_{diag}$ is a diagonal matrix with $\sigma_{ij} = \Sigma_{ij} = 0$ for $i \neq j$. HRP notably uses this result when recursively evaluating the portfolio weights subsequent to quasi-diagonalizing the covariance matrix. However, in finance it is often the case that instruments are not independent, and for this reason the inverse variance portfolio is far from ideal.

IMPROVED INVERSE APPROXIMATION

A notable improvement to the inverse variance portfolio can be found in López de Prado and Lewis [2018]. Specifically, were we to assume a constant correlation ρ between each instrument i, j with $i \neq j$, then $\sigma_{ij} = \rho\sigma_i\sigma_j$, and thus

$$\Sigma = (1 - \rho)\Sigma_{diag} + \rho\sigma\sigma^T,$$

where I is the identity matrix, and σ is the column vector of standard deviations. Given that this is simply a rank-one update to the matrix $(1-\rho)\Sigma_{diag}$, we can use the Sherman-Morrison Identity to evaluate its analytic inverse.

$$\Sigma^{-1} = \frac{1}{1-\rho}\Sigma_{diag}{}^{-1} + \frac{\rho}{(1-\rho)(1+\rho(N-1))}\left(\frac{1}{\sigma}\right)\left(\frac{1}{\sigma}\right)^T$$

In this equation $\left(\frac{1}{\sigma}\right)$ is, with slight abuse of notation, the vector with component i set to $\frac{1}{\sigma_i}$.

Observe that, were the covariance matrix to have this form, then the optimal minimum variance portfolio would take the form

$$w_i \propto \frac{1}{\sigma_i^2} + \frac{\rho \sum_j \frac{1}{\sigma_j}}{(1+\rho(N-1))\sigma_i},$$

while the optimal maximum sharpe ratio portfolio would take the form

$$w_i \propto \frac{\mu_i}{\sigma_i^2} + \frac{\rho \sum_j \frac{\mu_j}{\sigma_j}}{(1+\rho(N-1))\sigma_i}.$$

As one should expect, this reduces to the inverse variance portfolio in the case where $\rho = 0$. In practice, we can set ρ to equal the off-diagonal average correlation if we wish to approximate the covariance matrix this way.

PORTFOLIO ENHANCEMENT VIA RECURSIVE CLUSTERING

While the above enhancement may be a useful approximation in cases where the off-diagonal correlations are all similar, this clearly isn't the typical case. We must therefore reconsider use of this method to better leverage this result.

To facilitate discussion, let us first consider a portfolio of stocks that are exclusively in the Industrial or the Technology sector. In this scenario, we expect Industrial stocks to have returns similar to other Industrial stocks, and likewise Technology stocks returns similar to other Technology stocks. However, Industrial stock performance will be less similar to Technology stocks. Were we to sort the indices to place the Industrials first followed by the Technology stocks, we should expect the correlation matrix matrix to have a block-like formation. See Exhibit 1 for such an example of a hypothetical correlation matrix. Consider first the portfolio of just Industrials as portfolio 1, and the portfolio of just Technology stocks as portfolio 2. We could optimize portfolio 1, then do similarly for portfolio 2, and finally optimize this portfolio of optimized portfolios. While technically suboptimal, its performance would likely be near optimal.

This thought experiment brings up a few discussion questions:

1. Some Industrial stocks performs more similarly to Technology stocks than other Industrials, so why delineate by sector?
2. Within the Technology sector, stocks A and B may perform more similarly to one another and less like stock C, so why not further dissect the stocks within each sector?
3. Why limit the analysis to a portfolio of just Industrials and Technology stocks?

Thus, for a more general portfolio, we must generalize this optimization procedure

We refer to this generalized optimal portfolio procedure as Recursive Clustering Risk Parity (RCRP), which takes as inputs a covariance matrix of returns Σ and vector a, with $a = 1$

in the case of the minimum variance portfolio and $a = \mu$ in the max sharpe ratio portfolio. The procedure can be described as follows:

1. Form the correlation matrix ρ from the covariance matrix Σ for a certain set of instruments S.
2. Use advanced unsupervised learning techniques to cluster ρ and partition S into k disjoint sets of instruments S_i.
3. For each set S_i, extract the covariance matrix Σ_i and vector a_i restricted to those instruments
4. If Σ_i is too small for clustering, find the optimal portfolio vector w_i via the improve inverse portfolio optimization above using Σ_i and a_i. Otherwise, set $w_i = RCRP(\Sigma_i, a_i)$. Note that w_i has dimension $N_i = |S_i|$ and $1^T w_i = 1$.
5. Using w_i to define the change in basis elements, create the compressed covariance matrix Σ^{comp} and compressed vector a^{comp} for our portfolio of portfolios. Observe that $a_i^{comp} = a_i^T w_i$ $\Sigma_{ij}^{comp} = w_i^T \Sigma_{ij} w_j$, where Σ_{ij} is the submatrix of Σ with rows related to instruments in S_i and columns related to instruments in S_j. Find the cluster weights vector v via the improved inverse portfolio optimization using Σ^{comp} and a^{comp}. Note that v has dimension k, and $1^T v = 1$.
6. For $i = 1, ..., k$, set $w_i \rightarrow v_i * w_i$, then concatenate the weight vectors w_i into a single vector w, and return w. Note that $1^T w = 1$ as desired.

For step (2) in our procedure, we consider the clustering method formulated in López de Prado and Lewis [2018], though this can be enhanced should improved clustering techniques arise.

NUMERICAL EXAMPLES:

The code for this analysis, written in Python, can be found at: https://github.com/mlewis1729

A key takeaway is that there is significant improved performance to industry alternatives.

CONCLUSIONS

We have discussed a new procedure.

REFERENCES

López de Prado, Marcos, Building Diversified Portfolios that Outperform Out-of-Sample (May 23, 2016). Journal of Portfolio Management, 2016; https://doi.org/10.3905/jpm.2016.42.4.059. . Available at SSRN: https://ssrn.com/abstract=2708678 or http://dx.doi.org/10.2139/ssrn.2708678

López de Prado, Marcos and Lewis, Michael J., Detection of False Investment Strategies Using Unsupervised Learning Methods (August 18, 2018). Available at SSRN: https://ssrn.com/abstract=3167017 or http://dx.doi.org/10.2139/ssrn.3167017