‘Adaptive Seriational Risk Parity’ and other Extensions for Heuristic Portfolio Construction using Machine Learning and Graph Theory

Markus Jaeger[1*] Stephan Krügel[1†] Jochen Papenbrock[2‡] Peter Schwendner[3§]

[1] Munich Re Markets, Königinstraße 107, 80802 Munich, Germany
[2] Firamis GmbH, Robert-Kempner-Ring 27, 61440 Oberursel, Germany
[3] Zurich University of Applied Sciences, Institute of Wealth & Asset Management, Technoparkstrasse 2, CH-8401 Winterthur

Abstract

In this article, the authors present a conceptual framework named ‘Adaptive Seriational Risk Parity’ (ASRP) to extend Hierarchical Risk Parity (HRP) as an asset allocation heuristic. The first step of HRP (quasi-diagonalization) determining the hierarchy of assets is required for the actual allocation in the second step of HRP (recursive bisectioning). In the original HRP scheme, this hierarchy is found using the single-linkage hierarchical clustering of the correlation matrix, which is a static tree-based method. The authors of this paper compare the performance of the standard HRP with other static and also adaptive tree-based methods, but also seriation-based methods that do not rely on trees. Seriation is a broader concept allowing to reorder the rows or columns of a matrix to best express similarities between the elements. Each discussed variation leads to a different time series reflecting portfolio performance using a 20-year backtest of a multi-asset futures universe. An unsupervised representation learning based on this time series data creates a taxonomy that groups the strategies in high correspondence to the structure of the various types of ASRP. The performance analysis of the variations shows that most of the static tree-based alternatives of HRP outperform the single linkage clustering used in HRP on a risk-adjusted basis. Adaptive tree methods show mixed results and most generic seriation-based approaches underperform.

Keywords: Hierarchical Risk Parity, portfolio allocation, hierarchical structure, seriation

Introduction

To achieve more robust portfolios, there is a growing effort to replace mean-variance optimization in allocation models. Many modeling approaches reflect the hierarchical correlation dynamics of financial markets. Such approaches have gained a lot of attention in the academic and quantitative investment community and many extensions and modifications have been proposed in recent years.

Recent approaches for simulating realistic financial correlation matrices explicitly address hierarchy as stylized facts (see Huettner and Mai (2019), Gautier Marti (2019), Jaeger et al. (2021) and Papenbrock et al. (2021)). Modeling the correlation hierarchy of markets has also been used for the recognition of market regimes (Papenbrock and Schwendner (2015)).

Related approaches model market complexity as networks or use partitional/flat clustering. Examples are Onnela et al. (2003), Lohre, Papenbrock, and Poonia (2014), Baitinger and Papenbrock (2015) and Baitinger and Papenbrock (2017).

[*] majaeger@munichre.com
[†] skruegel@munichre.com
[‡] jp@firamis.de
[§] scwp@zhaw.ch

In this paper however, we focus on 'purely hierarchical' approaches. In recent years new algorithms for portfolio construction explicitly take hierarchy of financial markets into account, acknowledging the behavior of the financial markets similar to a complex system. The new methods address three major concerns of quadratic optimizers: instability, concentration and out-of-sample underperformance. Some of these hierarchical methods are even 'optimization free' recognizing traditional optimizers as vulnerable because they require a well-conditioned covariance matrix. These types of the sampling errors stemming from historical time series data are well documented e.g. in Pafka and Kondor (2002) and Menchero and Ji (2019).

According to Lopez de Prado (2016a) there are both numerical instabilities caused by noise (i.e. when the dimension of the covariance matrix outnumbers observations) and by signal (the source of this instability is distinct and unrelated to noise). Infinancial markets, correlation clusters exist as a consequence of hierarchical relationships de-stabilizing optimization.

Therefore, Lopez de Prado (2016b) proposes an optimization-free, heuristic algorithm called Hierarchical Risk Parity (HRP). It has been shown that such portfolios can outperform traditional approaches out-of-sample.

This paper starts with discussing the origins and properties of hierarchical approaches such as HRP. Second, we highlight practical implementation issues and discuss underlying assumptions. Third, we introduce variations, extensions and generalizations in a conceptual framework labeled "Adaptive Seriational Risk Parity" (ASRP).

In an empirical study we backtest the discussed ASRP variations using a multi-asset universe of 17 liquid futures markets. We further analyze the correlation hierarchy of the resulting portfolio return time series in a dendrogram and show their risk-adjusted returns relative to the HRP strategy. The risk-adjusted return measures are validated employing bootstrapping. Finally, we suggest to embed the discussed allocation variations in a "Triple AI" approach to validate investment strategies.

Origins of the Hierarchical Approaches

The idea of grouping stocks in hierarchical ways dates back to the 1960s (for a short introduction see Papenbrock (2011)). A seminal work on hierarchical clustering of financial markets was presented by Mantegna (1999), followed by more than two decades of research in correlations, hierarchies, networks and clustering in financial markets (see G. Marti et al. (2017) for a very comprehensive overview).

Some papers are based in cleaning or filtering the correlation or covariance matrix with a special focus on hierarchical clustering. Based on that input, in a second step a Global Minimum Variance Portfolio is built as for example in Tola et al. (2008)[1] and Bongiorno and Challet (2020). These approaches could be called Hierarchical Minimum Variance (HMV).

A second stream of literature employs factor models for identifying hierarchies: Tumminello, Lillo, and Mantegna (2007) and Avellaneda (2019).

This paper focuses on outright utilization of hierarchical clustering in portfolio allocation. Hierarchical models can build up portfolios bottom-up or top-down. A top-down example is the so-called 'Waterfall' approach by Papenbrock (2011) that was later further discussed in Raffinot (2017). It is a machine learning (ML) approach that uses a distance matrix based on a group of asset return time series and learns a hierarchical representations using a clustering procedure that visualized as a hierarchical tree or dendrogram. It reveals the hierarchical distance or proximity among assets in a first step.

Thereafter, it allocates capital according to the hierarchical splits from top to bottom in a waterfall-like style: at each binary split, it allocates capital 50/50 downwards to the leaves of the tree. We will call this procedure"In the following we will call it the"Hierarchical Equal Weight" (HEW) hereinafter.

The HEW approach implicitly assumes each splitting point in the cluster tree of two clusters on the left and right branch to exhibit the same risk level. Therefore, the splitting of capital notional at each consecutive branch in the top-down hierarchy assumes risk parity among the two branch clusters.

[1]Markowitz with Average and Single Linkage as only these create positive semi-definite matrices.

These approaches can be varied in a number of ways: Different investors may choose different linkage types of the clustering algorithms representing differently implemented strategies. For example, a single linkage hierarchical clustering exhibits a chaining effect whereas a Ward-based hierarchical clustering tries to create clusters of similar sizes (Ward Jr (1963)) . The results are based on the correlation coefficient so that individual risk measures like variance or reward measures (e.g. expected returns) may enter the model. The distance matrix could incorporate correlation and risk level information of the assets.

Lopez de Prado (2016b) proposed as similar method called 'Hierarchical Risk Parity' (HRP). The author introduces it in the following way: 'the Hierarchical Risk Parity approach addresses three major concerns of quadratic optimizers, in general, and Markowitz's critical line algorithm (CLA), in particular: instability, concentration, and underperformance. HRP applies modern mathematics (graph theory and machine-learning techniques) to build a diversified portfolio based on the information contained in the covariance matrix. However, unlike quadratic optimizers, HRP does not require the invertibility of the covariance matrix. In fact, HRP can compute a portfolio on an ill-degenerated or even a singular covariance matrix—an impossible feat for quadratic optimizers. Monte Carlo experiments show that HRP delivers lower out-of-sample variance than CLA, even though minimum variance is CLA's optimization objective. HRP also produces less risky portfolios out of sample compared to traditional risk parity methods.' HRP also asserts that the correlation structure contains ordinal information, which can be exploited by organizing the assets into a hierarchy. The goal of Hierarchical Risk Parity is to translate/reorganize the covariance matrix such that it is as close as possible to a diagonal matrix, without altering the covariance estimates. The minimum variance portfolio of a diagonal matrix is the inverse variance portfolio. For this reason, the method is also known as 'Hierarchical Minimum Variance.'

HRP basically consists of two steps:

1. 'quasi-diagonalization': permutation of the covariance matrix according to a hierarchical clustering approach, based on hierarchical clustering (tree/dendrogram)
2. 'recursive bisectioning': splitting the matrix in equally sized clusters and allocating capital in inverse proportion to the risk of the clusters. This splitting in equally sized clusters is executed recursively until the clusters consist of single assets. So the algorithm naively bisects all assets into two equal-sized groups. The covariance matrix is simply split into two equal-sized covariance submatrices.

HRP incorporates similarities and differences to HEW: they both start with a hierarchical clustering and distribute capital according to binary splittings. Also, both may be adapted by changing the hierarchical clustering and/or distance matrix. The first difference between the approaches is that HRP divides the matrix in equally sized bins whereas the HEW approach splits along the hierarchical tree structure which is intuitive. The second difference is that the HEW approach creates a 50/50 notional split whereas the HRP splits according to the inverse risk of the two clusters. If the tree is symmetric and contains clusters of equal risk at all tree levels, then HRP and HEW do actually coincide. A final difference is that HRP uses both the correlation matrix and the volatilities of the assets to weight the portfolio. In contrast to HRP, HEW only uses all pairwise correlation information. We will show however that HEW can be modified to also account for the volatilities of the assets.

HRP exhibits very robust results out-of sample in many data/market situations. It minimizes the variance but not in a too concentrated way as can be observed with the Minimum Variance approach. Its weight structure is balanced as can be quantified by a number of diversification measures. In this way it is relatively prone to idiosyncratic and systemic shocks.

Agony of Choice

HRP and related approaches based on hierarchical clusters can be designed, configured and parameterized on an almost infinite space of combinations and variations. The reported three steps of HRP may be adapted as often done in the literature. For example, the tree clustering step can be done with different linkage approaches of hierarchical clustering. The sectioning step does not necessarily have to be carried out by equal size splits but rather explicitly using the dendrogram structure. The bisectioning might also be stopped at a certain point, especially when portfolios are large, when correlations increase and clustering becomes less separated. In these cases, cutting of the dendrogram may take place at any plausible point, e.g. when an optimal amount of clusters is reached or when the number of clusters approximates some external grouping criteria like industry sector, style, geography or (sub) asset class.

Consequently, there are many more choices for preprocessing regarding the input matrices of correlation, similarity, and distance. Next the risk of an asset or cluster may be estimated by variance, volatility or other risk measures and the correlations within and across clusters may be included. Tail correlation and other higher-order effects may also be modeled.

Next there are configuration choices related to the investment mandate and to the institutional requirements and constraints addressing questions such as: how large is the universe; what is the rebalancing and data frequency; what is the trading cost; which amount of turnover is allowed during the year; how concentrated can a portfolio be, how is it supposed to respond to specific and systematic shocks, are there group/box/turnover constraints?

Introducing more or less sophistication and precision in modeling leads to different sets of assumptions and potential over-expressions in a balance between precision, bias and time lag. For example, a higher update frequency for a sampled covariance matrix might decrease the tracking error of a portfolio while at the same time increase costly portfolio turnover. The portfolio should reflect the data generating process, the non-stationarity, the asset class and hedging relationships as well as the market conditions.

These examples show that finding the ‘right’ HRP configuration or hierarchical cluster model is not trivial. Therefore, the following chapter gives an overview of the variety of HRP variations and some examples on how to address these issues.

Literature on HRP Extensions

Tree-based Sectioning

In the standard HRP approach, bisectioning might separate highly correlated assets. The ‘Waterfall’ approach of HEW by Papenbrock (2011) introduces a splitting according to (correlation cluster) tree structure, therefore called a ‘tree-based’ sectioning. In contrast, the standard bisection step of HRP might separate highly correlated assets erroneously. The loss of information resulting from ignoring the correlation between two clusters, however, should be minimized.

The next step in HRP weights the two split clusters according to the inverse of their risk - in contrast, HEW weights them equally.

Other approaches like Alipour and Rounds (2016) and Lohre, Rother, and Schäfer (2020) also apply the tree-sectioning method. They are thereby similar to the HEW approach. However, they use alternative tree construction methodologies and not equally weight the clusters at each sectioning step.

Pfitzinger and Katzke (2019) developed a flexible extension of the bisectioning step of HRP where there is a parameter tau between zero and one, parameterizing a bisection of HRP at one end and the split according to the tree-section method on the other.

The tree-section step seems to be the superior and a more intuitive way to approach the recursive sectioning step of HRP. However, note that such methods not only rely on tree cluster quality for the diagonalization step but also for the sectioning. Whenever there is a suboptimal fit of a tree cluster structures to real data, the tree-section strategies are subject to higher model risk.[2]

[2] The authors of https://www.linkedin.com/pulse/shortfalls-hierarchical-risk-parity-rafael-nicolas-fermin-cota/ additonally report: ‘A second shortfall we have identified occurs when there is a highly correlated group of assets. We have noticed poor performance out-of-sample compared with a naive equal-weight or inverse volatility weighting. To address this shortfall, we have introduced an early stopping step within the recursive bisection that identifies when a submatrix of the covariance matrix contains high pairwise correlations. Weighting the assets according to equal-weight or inverse volatility has proven to provide better out-of-sample results based on our research.’

Cleaned/Filtered Correlation/Covariance Matrix

Molyboga (2020) introduce a "Modified HRP" (MHRP) approach that 'extends the HRP approach by incorporating three intuitive elements commonly used by practitioners. The new approach (1) replaces the sample covariance matrix with an exponentially weighted covariance matrix with Ledoit–Wolf shrinkage; (2) improves diversification across portfolio constituents both within and across clusters by relying on an equal volatility, rather than an inverse variance, allocation approach; and (3) improves diversification across time by applying volatility targeting to portfolios. The author examines the impact of the enhancements on portfolios of commodity trading advisors within a large-scale Monte Carlo simulation framework that accounts for the realistic constraints of institutional investors. The author finds a striking improvement in the out-of-sample Sharpe ratio of 50%, on average, along with a reduction in downside risk.'

Jothimani and Bener (2019) combine the idea of HRP with the robust Gerber statistics (HRP-GS). They test the model using stocks comprising the TSX index for a time period of 10 years ranging from 2007 to 2016. Results suggest that the proposed HRP-GS model outperforms the standard HRP model.

Alternative Codependence and Distance Metrics

Barziy and Chlebus (2020) compare the HRP performance under various codependence and distance metrics. The algorithm is tested using the modifications of codependence metrics of the instruments in a portfolio (distance correlation, mutual information, variation of information), and distance metrics to transform the codependence matrix into the distance matrix (angular, absolute angular and squared angular).

Jain and Jain (2019) highlight the need to account for covariance misspecification and test for predictive ability on out-of-sample portfolio performance. Next to sample-based covariance (SMPL) they use exponentially weighted moving average (EWMA) and dynamic conditional correlation GARCH (DCC-GARCH). They find that when the covariance estimates are crude, inverse volatility weighted portfolios are more robust, followed by HRP. Minimum variance and maximum diversification are most sensitive to covariance misspecification. HRP seems to offer a compromise; it is less sensitive to covariance misspecification when compared with minimum variance or maximum diversification portfolio, while it is not as robust as the inverse volatility weighted portfolio.

Lohre, Rother, and Schäfer (2020) test a codependence measure incorporating tail behaviour of assets. In a multi-factor multi-style universe they investigate the Tail-HRP approach using the CSR-estimator (Conditional Spearman's Rho) by Schmid and Schmidt (2007). Also, they use the following distance measure: d(X, Y) = -log(lambda) where lambda is the tail dependence coefficient.

Simultaneous Modifications to HRP

Pfitzinger and Katzke (2019) introduce a constrained HRP with box and group constraints relevant for solving practical portfolio problems. These constraints can be combined with the other extensions that the authors describe: follow more a bisectioning or a tree sectioning (with respective parameter tau) plus an alternative quasi-diagonalization generated by 'genetic HRP.'

Raffinot (2018) combines several potential modifications of HRP in the HERC algorithm: the cluster dendrogram can be pruned at some point corresponding to some optimality criterion regarding the number of flat/partitional clusters and cluster risk contributions can be measured by variance, standard deviation, conditional value at risk (CVaR) and conditional drawdown at risk (CDaR). Different weighting schemes can be applied to the assets within the clusters.

Other HRP Solutions

Burggraf (2020) employs HRP to a large portfolio of cryptocurrencies. An out-of-sample comparison with traditional risk-minimization methods reveals that HRP outperforms in terms of tail risk-adjusted return, thereby working as a potential risk management tool that can help cryptocurrency investors to better manage portfolio risk.

Finally, HRP leads occasionally to higher turnover as reported in Kolrep et al. (2020). The authors suggest a 'Smooth HRP' method to mitigate this issue.

Adaptive Seriational Strategies in Heuristic Portfolio Construction

The main idea of "Adaptive Seriational Risk Parity" (ASRP) is that the tree-based quasi-diagonalization step of HRP is adaptively extended to a wider class of methodologies to diagonalize a distance matrix based on a technique called seriation. Remember the goal of HRP to be translating/reorganizing the covariance matrix such that it is as close as possible to a diagonal matrix, without altering the covariance estimates. The minimum variance portfolio of a diagonal matrix is the inverse variance portfolio. For this reason, HRP is sometimes also called 'Hierarchical Minimum Variance.' Inverse-variance asset allocation is most appropriate for assets with an approximately diagonal correlation matrix. High correlations are placed adjacently and close to the matrix diagonal, achieving the desired quasi-diagonal structure.

There are many ways to diagonalize a matrix so it does not necessarily require a tree structure. The tree structure however often has the advantage of fast computation and it extracts the hierarchical nature of complex data. However, sometimes non-tree-based seriation can be more appropriate.

Even when sticking to the tree structures there are numerous hierarchical clustering algorithms and not just a single one like 'single linkage' as in standard HRP. Conclusively, there are many tree-based and non-tree-based seriation methods. Therefore, we introduce an automated selection procedure that picks the most suitable seriation for the quasi-diagonalization step in HRP.

As markets evolve over time the choice of seriation alters, requiring updating as whenever another seriation method becomes more appropriate. Consequently, the seriation method could be chosen adaptively depending on market situations.

Even if there is no adaptive procedure needed, it is crucial to decide which seriation method to use for an entire strategy or data set as a one time choice.

The following section will describe methodologies for choosing appropriate seriation and criteria. Both seriation and criteria may or may not be based on trees.

Exhibit 1 shows our ASRP construction hierarchy of strategies. The first choice is between static seriation-based and tree-based approaches. Tree-based approaches include static and adaptive variations. HRP uses single linkage as an example for a static tree-based method. Adaptive methods may be based on the distance matrix, but also on other methods as discussed above. We compare this construction hierarchy to the empirical hierarchy of performance similarity for the empirical dataset.

Empirical Study

We rely on a multi-asset universe of equity index, sovereign bond and commodity futures from 2000-05-03 to 2020-06-30 with daily frequency as in Jaeger et al. (2021) and Papenbrock et al. (2021). Exhibit 2 shows the Bloomberg tickers, asset classes, currency and names of the 17 futures markets. At each monthly rebalancing date, the ASRP variations are applied to a rolling historical one-year window. The leverage of the resulting portfolio allocation is set with the aim to realize a 5% volatility target. This is achieved by setting the leverage to the ratio of the target volatility to the maximum of the empirical portfolio volatilities computed in rolling windows of 20 and 60 trading days. We account for 2 bp transaction cost for a half-trip.

Exhibit 2: Futures Universe

Ticker	Asset class	Currency	Name
CLA Comdty	Commodities	USD	NYMEX WTI Light Sweet Crude Oil
GCA Comdty	Commodities	USD	COMEX Gold
SIA Comdty	Commodities	USD	COMEX Silver
BZA Index	Equities	BRL	BM&F IBOVERSPA
ESA Index	Equities	USD	CME E-mini S&P 500

Ticker	Asset class	Currency	Name
HIA Index	Equities	HKD	HKFE Hang Seng
NKA Index	Equities	JPY	OSE Nikkei 225
NQA Index	Equities	USD	CME E-mini NASDAQ-100
SMA Index	Equities	CHF	Eurex SMI
VGA Index	Equities	EUR	Eurex EURO STOXX 50
XPA Index	Equities	AUD	ASX SPI 200
Z A Index	Equities	GBP	ICE FTSE 100
CNA Comdty	Fixed Income	CAD	10Y Canadian GB
G A Comdty	Fixed Income	GBP	ICE Long Gilt
RXA Comdty	Fixed Income	EUR	Eurex 10Y Euro-Bund
TYA Comdty	Fixed Income	USD	CBOT 10Y US T-Note
XMA Comdty	Fixed Income	AUD	ASX 10Y Australian T-Bonds

The Impact of Quasi-Diagonalization

We use this multi-asset universe and take 2019 market data for estimation. The period from 2020-01-01 to 2020-06-30 is the investment period for an out-of-sample test. The standard HRP strategy with static weights from 2020-01-01 without further rebalancing and without a volatility target results in a Sharpe ratio of 0.732 for the six-month investment period.

To assess the impact of the quasi-diagonalization step, we sample 10.000 random asset permutations for the quasi-diagonalization step instead of the single-linkage hierarchical clustering of HRP. Thereafter, we apply the bisection step of the standard HRP strategy to each permutation using the original covariance matrix, measure the annualized Sharpe ratios (SR) and get the following results:

Exhibit 3: Statistics of Permutations of Quasi-Diagonalization Step

Stats	Sharpe Ratio
Min.	0.070
1st Qu.	0.629
Mean	0.747
Median	0.755
3rd Qu.	0.871
Max.	1.364
Stddev	0.184
HRP	0.732

Exhibit 4 shows the smoothed density of Sharpe ratios resulting from HRP strategies across the sampled random matrix permutations. The Sharpe ratio of the unchanged HRP strategy (0.732) is marked as a vertical red line.

The large standard deviation $\sigma = 0.184$ of Sharpe ratios translates into a half-width of about $\sigma \times 2.4 = 0.44$ illustrating the permutation in this data set to play a huge role. The quasi-diagonalization step of HRP has consequences potentially also in other data sets and set-ups.

Seriation-based Quasi-Diagonalization

Seriation, also referred to as ordination or matrix permutation, dating back to (Petrie (1899)). One more recent application of seriation is the visualization of tables, matrices, clusters and networks outlined in Behrisch et al. (2016).

In this paper we focus on the criteria and seriation methods as described in Hahsler, Hornik, and Buchta

Exhibit 1: Hierarchical Construction Concept of Adaptive Seriational Hierarchical Risk Parity

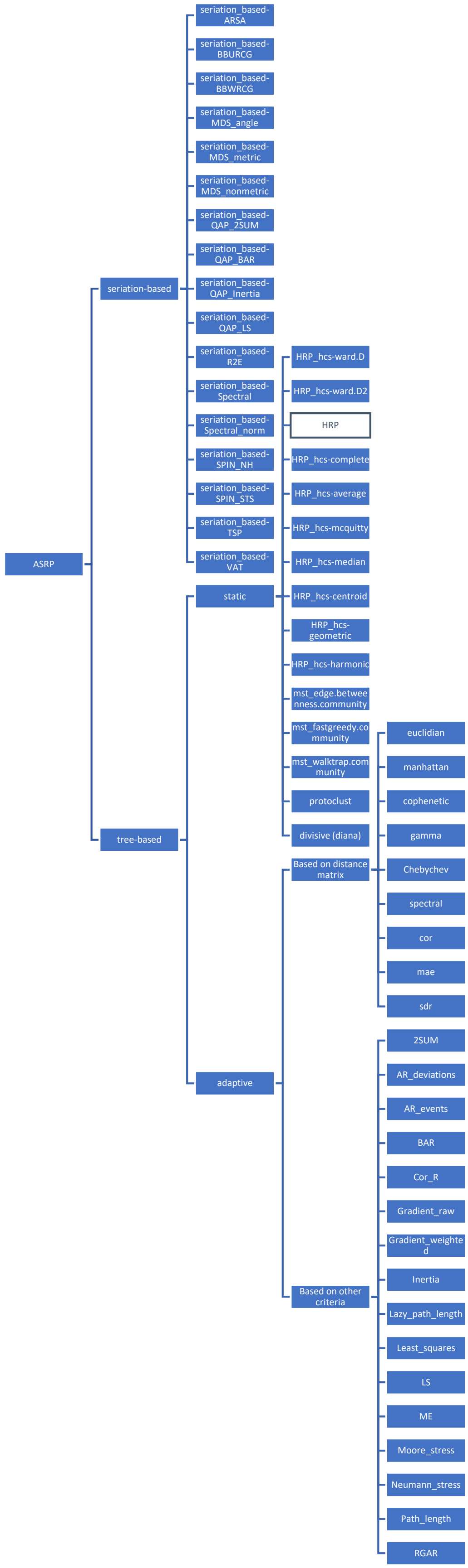
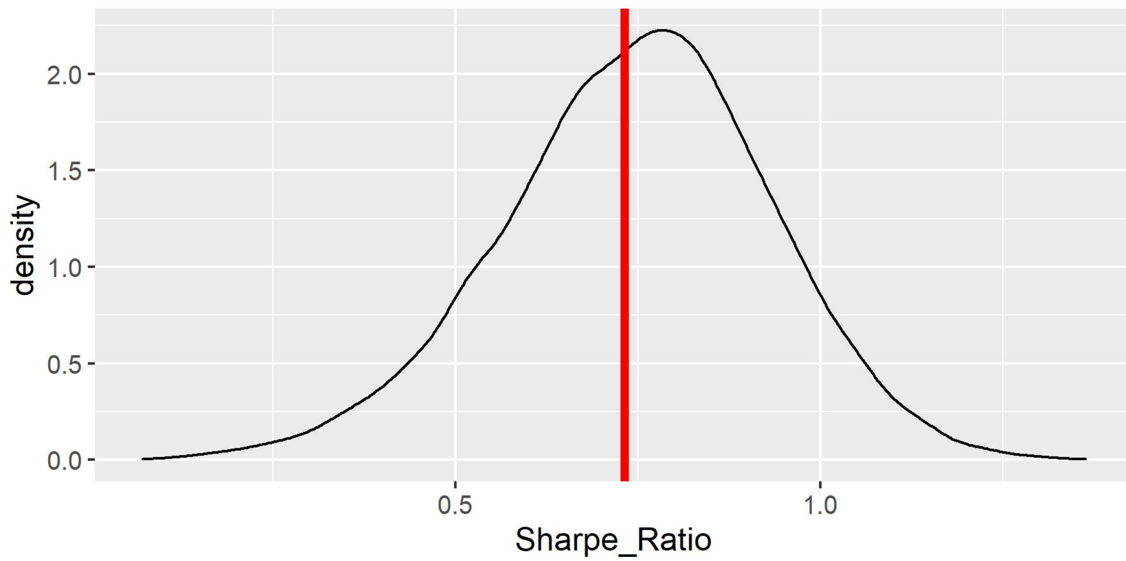


Exhibit 4: Density of Sharpe Ratios from random permutations of correlation matrix in quasi-diagonalization step of HRP



(2008) and Hahsler (2017). These resources contain a number of seriation methods (tree and non-tree) as well as criteria that measure the quality of some matrix permutations.

One such seriation method called 'Inertia' is further described in Hahsler, Hornik, and Buchta (2008) and Caraux and Pinloche (2004). Alipour and Rounds (2016) use a variation of it for HRP-like strategies. Quantum computers could solve this problem but currently it can also be addressed by a genetic algorithm as shown in Pfitzinger and Katzke (2019) if the number of iterations is reasonably small.

Specifically, Alipour and Rounds (2016) use a variation of 'Inertia' to quasi-diagonalize the matrix for the first HRP step. Many alternatives to this seriation are outlined in chapter chapter 'Non-tree-based quasi-diagonalization.'

After the inertia-like seriation step the authors do not execute the bisection step of HRP but insert another additional step that produces a tree based on the seriation. This in turn enables fine-tuned sectioning that better addresses a block structure in the diagonalization, namely a tree-sectioning instead of a naive bisectioning[3]. According to the authors, this approach yield more robust results than HRP. An explanation could be that the diagonalization of the seriation method is very effective and that the tree construction based on the seriation enables a tree-sectioning instead of naive bisectioning. An important step would be to analyze the performance attribution of 1) the introduction of the inertia-like seriation and 2) the tree-sectioning instead of bisectioning.

Summarizing, the methodology of Alipour and Rounds (2016) can be generalized in the following two ways:

1. The seriation method may be replaced by numerous alternatives, potentially better meeting some criteria for quasi-diagonalization
2. The extra step of creating a tree out of a seriation may be applied or skipped, depending on the extra performance contribution of this step.

Step 1 and 2 in combination are very advantageous for HRP-like portfolio construction as first inverse-variance asset allocation is most appropriate for assets with an approximately diagonal correlation matrix and second the potential block structure is recognized by the tree-sectioning and not ignored as in the naive bisectioning.

[3] It is a recursively searching over the range of potential split position that optimizes a suitably chosen metric, such as the mean absolute distance matrix values of the off-diagonal blocks' entries or the mean absolute correlation of off-diagonal cluster blocks in the matrix.

However, superimposed tree structures for tree-sectioning may carry additional model risk in that the tree might not represent the structures very well.

In our empirical study we use the non-tree-based (non-hierarchical) seriation methods described in Hahsler, Hornik, and Buchta (2008) and shown in Exhibit 5, accepting a distance matrix as input.

Exhibit 5: Seriation Methods

Name	Seriation Type
ARSA	Anti-Robinson seriation by simulated annealing
BBURCG	Anti-Robinson seriation by branch-and-bound (unweighted gradient measure)
BBWRCG	Anti-Robinson seriation by branch-and-bound (weighted gradient measure)
MDS_angle	Multidimensional scaling (angle)
MDS_metric	Multidimensional scaling (metric)
MDS_nonmetric	Multidimensional scaling (non-metric)
QAP_2SUM	Quadratic assignment problem formulation (2-Sum Problem)
QAP_BAR	Quadratic assignment problem formulation (banded anti-Robinson form)
QAP_Inertia	Quadratic assignment problem formulation (inertia criterion)
QAP_LS	Quadratic assignment problem formulation (Linear Seriation Problem)
R2E	Rank-two ellipse seriation
Spectral	Spectral seriation
Spectral_norm	Spectral seriation normalized L
SPIN_NH	Sorting Points Into Neighborhoods (neighborhood algorithm)
SPIN_STS	Sorting Points Into Neighborhoods (Side-to-Side algorithm)
TSP	Traveling salesperson problem
VAT	Visual Assessment of Clustering Tendency.

The VAT method by Bezdek and Hathaway (2002) creates an order based on Prim's algorithm for finding a minimum spanning tree in a weighted connected graph representing the distance matrix. The order is given by the order in which the nodes (objects) are added to the MST.

Tree-based Quasi-Diagonalization

The topology or inherent shape and form of an object is important. A formal definition of hierarchical structure is provided by ultrametric topology. According to Murtagh (2007) ultrametricity is a pervasive property of observational data. That is why identifying and exploiting ultrametricity is important when analyzing complex financial data. Ultrametricity is a natural property of sparse, high-dimensional spaces and it emerges as a consequence of randomness and the law of large numbers. The 'strong triangular inequality' or ultrametric inequality is: $d(x, z) \leq max[d(x, y), d(y, z)]$ for any triplet x, y, z. The subdominant ultrametric is also known as the ultrametric distance resulting from the single linkage agglomerative hierarchical clustering method as used in HRP. Closely related graph structures include the minimum spanning tree. Agglomerative nesting is a bottom-up procedure where objects initially represent individual clusters, and are successively merged into larger clusters until the full hierarchical structure is obtained. Single linkage clustering is a particular type of agglomerative nesting which calculates the distance between two clusters as the shortest distance between any member of one cluster to any member of another cluster. By contrast, divisive analysis is a top-down approach which begins with a single cluster containing all objects and successively sub-divides assets into smaller clusters until each cluster contains only a single object. Characterization, stability and convergence of hierarchical clustering can for example be studied in Carlsson and Mémoli (2010).

The subdominance provides a good fit to a given distance but it suffers from the 'friends of friends' or chaining effect. Many other hierarchical clusterings approaches have been developed with specific properties like creating spherical and more equally sized clusters as in the Ward clustering (Ward Jr (1963)).

Tree-based quasi-diagonalization always has the advantage that most of them are fast to compute and they extract block structures which are helpful for the tree-sectioning step in HRP. They are not so much optimized for diagonalization - the first step of HRP.

HRP-style portfolio construction exhibits further advancements in the future as hierarchical clustering and exploitation of ultrametric seem to be an ongoing, active research field with applications in many domains. Examples are hierarchical clustering with prior knowledge (e.g. Ma and Dhavala (2018)) and ultrametric fitting (e.g. Chierchia and Perret (2019)).

Recently, Prado (2019) introduced an approach based on "theory-implied correlation matrices" (TIC) approach to estimate forward-looking correlation matrices implied by economic theory. Given a particular theoretical representation of the hierarchical structure that governs a universe of securities, the method fits the correlation matrix complying with that theoretical representation of the future. The output is a tree so it is straightforward to use it in the quasi-diagonalization step of HRP. Babynin (2020) illustrates this idea.

We provide a description of the hierarchical cluster methods used in the empirical part of this paper:

- Fast hierarchical, agglomerative clustering routines are available in 'fastcluster' (Müllner (2013), R-version 1.1.25, also available in Python)
- another approach is to first build a minimal spanning tree with 'igraph' and then find a hierarchical community (Csardi and Nepusz (2006), R-version 1.2.5, also available in Python)
- minimax linkage hierarchical clustering is available in 'protoclust'(Bien and Tibshirani (2011), R-version 1.6.3)
- divisive hierarchical clustering is available in 'cluster' (Kaufman and Rousseeuw (1990), R-version 2.1.0)
- further agglomerative methods and dendrogram descriptive measures are available in 'mdendro' (Fernández and Gómez (2008), R-version 1.0.1)

Exhibit 6 shows names and the type of the clustering approaches used:

Exhibit 6: Clustering Methods

Name	cluster type
ward.D	agglomerative
ward.D2	agglomerative
single	agglomerative
complete	agglomerative
average	agglomerative
mcquitty	agglomerative
median	agglomerative
centroid	agglomerative
mst_edge.betweenness.community	hierarchical graph community
mst_fastgreedy.community	hierarchical graph community
mst_walktrap.community	hierarchical graph community
protoclust	protoclust
diana	divisive
geometric	agglomerative
harmonic	agglomerative

The following two sections define criteria on how to find the 'best' clustering approach. The first paragraph compares the original correlation distance matrix with the hierarchical clustering output. The second paragraph introduces criteria for the quality of seriations (tree-based seriations in our case).

Strategy names are given by 'HRP_hcs_XXX' with the following meanings:

- hcs: hierarchical clusterings
- XXX: one of the clustering names form the table

Adaptive Tree-based Strategies Based on the Distance Matrix

The quasi-diagonalization step reorders the rows and columns in such a way that the largest values lie close to the diagonal. This is achieved by rearranging the matrix based on the ordering generated by the cluster algorithm as described in the previous chapter.

A criterion to find the best match between a hierarchical cluster and a given distance matrix is to generate the ultametric distance of a hierarchical cluster and compare it with the original distance matrix without the clustering.

The cophenetic distance between two observations that have been clustered is defined to be the intergroup dissimilarity at which the two observations are first combined into a single cluster. In a dendrogram this can be compared by traversing from one leaf to another and recording at which dendrogram height the two leafs are connected.

To derive the ultrametric distance matrix, we first compute all pairwise cophenetic distances and then order the matrix identical to the original distance matrix. Ultrametric matrices of different clustering methods can be compared to each other to find out how 'close' or 'similar' the methods are. Also, the ultrametric matrix can be compared to the original distance matrix in the same way.

We use the following measures (some of them from package 'clue,' R-version 0.3-57, see Hornik (2005)) with the following definitions

u: ultrametric distance matrix v: original distance matrix

Exhibit 7: Distance Measures

Name	description
euclidean	square root of the sum of the squared differences of u and v
manhattan	sum of the absolute differences of u and v
cophenetic	product-moment correlation of u and v
gamma	rate of inversions between u and v
Chebyshev	maximum of the absolute differences of u and v
spectral	spectral norm (2-norm) of the differences of u and v
mae	normalized mean absolute error
sdr	range of the maximum and minimum ultrametric distances, divided by the range of the maximum and minimum original distances

Strategy names are given by 'HRP_hcs-adaptive_XXX_YYY' with the following meanings:

- hcs-adaptive: adaptive hierarchical clusterings
- YYY: criterion from the table above
- XXX: either from the clue or mdendro package

Adaptive Tree-based Strategies Based on Other Criteria

Based on a given distance matrix, the effectiveness of a permutation can be evaluated by certain criteria. Exhibit 8 shows our selected seriation criteria as described in Hahsler, Hornik, and Buchta (2008) and Hahsler (2017).

Exhibit 8: Seriation Criteria

Name	description
2SUM	2-Sum criterion
AR_deviations	Anti-Robinson deviations
AR_events	Anti-Robinson events
BAR	Banded anti-Robinson form (BAR)

Name	description
Cor_R	Measure of Effectiveness for the Moment Ordering Algorithm
Gradient_raw	Gradient measure (raw)
Gradient_weighted	Gradient measure (weighted)
Inertia	Hamiltonian path length
Lazy_path_length	Inertia criterion
Least_squares	Least squares criterion
LS	Linear seriation criterion (LS)
ME	Measure of effectiveness
Moore_stress	Stress (Moore neighborhood)
Neumann_stress	Stress (Neumann neighborhood)
Path_length	Hamiltonian path length (PL)
RGAR	Relative generalized Anti-Robinson events

Strategy names are given by 'HRP_hcs-adaptive_criterions_XXX_YYY' with the following meanings:

- hcs-adaptive: adaptive hierarchical clusterings
- criterions: Adaptive tree-based strategies based on other criteria
- XXX: criterion from the table above
- YYY: criterion is minimized or maximized

Backtests of all ASRP Variations

We compute backtests of all 57 ASRP variations applied to the multi-asset futures portfolio from 2000-05-03 to 2020-06-30 and compute the 57x57 correlation matrix between all strategy return time series. Exhibit 9 shows the result of a single-linkage clustering applied to this correlation matrix and the related Sharpe Ratios of each strategy relative to HRP.

We compare the strategy correlation hierarchy from the dendrogram to the construction hierarchy of Exhibit 1. Most seriation-based strategies are in cluster 1. Static tree-based strategies can be found in clusters 2, 3 and 4. Adaptive tree-based strategies are in all clusters except cluster 1. This confirms our initial hypothesis from the broad Sharpe ratio density in Exhibit 3 about the importance of the quasi-diagonalization step.

The right hand side of the dendrogram shows the relative Sharpe ratios to HRP for each strategy. Most seriation-based strategies seem to underperform HRP on a risk-adjusted basis. Most static tree-based variations outperform HRP, whereas adaptive tree-based methods show mixed results.

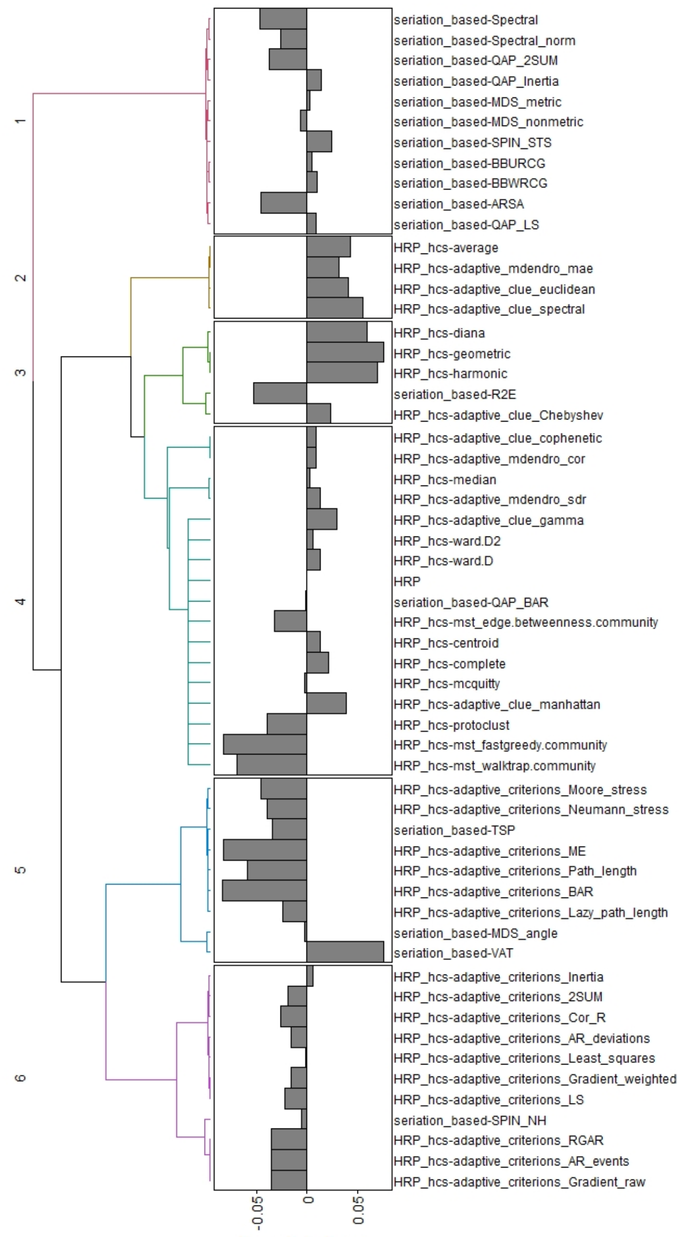
To assess the robustness of the dendrogram, we employ a bootstrap study: 1000 resamples of the asset return time series with a block length of 60 days lead to 1000 portfolio return time series. Across the lower triangle of the 57x57 strategy correlation matrix, we compute the ratios of all matrix elements of the point estimate of correlation divided by the standard deviation of correlations across the 1000 resamples. The minimum signal/noise ratio across all matrix elements of the lower triangle is 161. This large number points to only a small statistical correlation noise.

Conclusion

We present a systematic approach to generate a family of variations of the classical Hierarchical Risk Parity (HRP) for portfolio construction and asset allocation. These variations are directed towards the first step of HRP (quasi-diagonalization). We call this concept "Adaptive Seriational Risk Parity" (ASRP). A randomization of the quasi-diagonalization step shows a large variation of Sharpe Ratios and thus motivates the large impact of the model choice in this step on the resulting performance.

Backtests of all resulting 57 strategies with a multi-asset futures universe of 17 liquid markets across 20 years of data leads to a taxonomy-like representation of the resulting 57 strategy return time series in the form of a dendrogram. A bootstrap validation points to the robustness of this dendrogram. The pronounced hierarchy

Exhibit 9: Hierarchy of portfolio return correlations across ASRP strategies together with their Sharpe Ratios relative to HRP



of strategies almost resembled the original construction hierarchy of ASRP and thus confirms the large impact of the quasi-diagonalization step in HRP-style strategies. From the viewpoint of risk-adjusted returns, most of the static tree-based alternatives of HRP outperform, whereas seriation-based and adaptive methods tend to underperform with a few notable exceptions like the seriation-based VAT and SPIN_STS strategies.

Outlook: Synthetic Data and Explainable Machine Learning

The overall ambition to develop robust and transparent investment strategies requires several additional building blocks based on intelligent analytics.Together with the procedure to compute synthetic correlation matrices in Papenbrock et al. (2021) and the procedure to deliver local and global explanations using the SHAP framework discussed in Jaeger et al. (2021), this paper addresses variations of the allocation step. Thus it complements the resulting workflow of a 'Triple AI approach' (AI: analytical intelligence) for robust, adaptive, risk-based portfolio construction. It can be used to test, explore and understand the variations of HRP described before and consists of the following three steps:

1. a market generator to create synthetic time series or synthetic correlation data for Monte Carlo simulation, both serving as appropriate input to the HRP variations. Examples are 'Matrix Evolutions' (Papenbrock et al. (2021)) and CorrGAN (Gautier Marti (2019))
2. one or several of the ASRP alternatives to HRP as presented in this paper (some of them use tree-based representation learning)
3. 'explainable machine learning' to explain and understand the risk and performance attributions of single strategies and the drivers of a performance ranking in terms of features of the synthetic time series (as described in Jaeger et al. (2021) and Papenbrock et al. (2021)).

Acknowledgement

The implementation was sponsored by Munich Re Markets. We appreciate the infrastructure by Open Telekom Cloud and the NVIDIA GPU resources provided for this research.

We also acknowledge support from the European Union's Horizon 2020 research and innovation program "FIN-TECH: A Financial supervision and Technology compliance training programme" under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA) and from the COST Action CA19130 "Fintech and Artificial Intelligence in Finance - Towards a transparent financial industry."

References

Alipour, Clemens Adolphs, Elham, and Maxwell Rounds. 2016. "Quantum-Inspired Hierarchical Risk Parity." *1QBit*.

Avellaneda, Marco. 2019. "Hierarchical PCA and Applications to Portfolio Management." http://arxiv.org/abs/1910.02310.

Babynin, Andrey. 2020. "Hierarchical Risk Parity in Portfolio Construction." *Medium*. https://medium.datadriveninvestor.com/hierarchical-risk-parity-in-portfolio-construction-fc368db18c78.

Baitinger, Eduard, and Jochen Papenbrock. 2015. "Interconnectedness Risk and Active Portfolio Management: The Information-Theoretic Perspective." *Journal Of Network Theory In Finance* 1 (4): 1–25. https://doi.org/10.21314/JNTF.2015.012.

———. 2017. "Interconnectedness Risk and Active Portfolio Management." *Journal of Investment Strategies* 6 (2): 63–90. https://doi.org/10.21314/JOIS.2017.081.

Barziy, Illya, and Marcin Chlebus. 2020. "HRP performance comparison in portfolio optimization under various codependence and distance metrics." Working Papers 2020-21. Faculty of Economic Sciences, University of Warsaw. https://ideas.repec.org/p/war/wpaper/2020-21.html.

Behrisch, Michael, Benjamin Bach, Nathalie Henry Riche, Tobias Schreck, and Jean-Daniel Fekete. 2016. "Matrix Reordering Methods for Table and Network Visualization." *Computer Graphics Forum* 35 (3): 693–716. https://doi.org/10.1111/cgf.12935.

Bezdek, James C, and Richard J Hathaway. 2002. "VAT: A Tool for Visual Assessment of (Cluster) Tendency." In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02ch37290)*, 3:2225–30. IEEE.

Bien, Jacob, and Robert Tibshirani. 2011. "Hierarchical Clustering with Prototypes via Minimax Linkage." *Journal of the American Statistical Association* 106 (495): 1075–84. https://doi.org/10.1198/jasa.2011.tm10183.

Bongiorno, Christian, and Damien Challet. 2020. "Reactive Global Minimum Variance Portfolios with $k-$BAHC Covariance Cleaning."

Burggraf, Tobias. 2020. "Beyond Risk Parity a Machine Learning-Based Hierarchical Risk Parity Approach on Cryptocurrencies." *Finance Research Letters*, April, 101523. https://doi.org/10.1016/j.frl.2020.101523.

Caraux, Gilles, and Sylvie Pinloche. 2004. "PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order." *Bioinformatics* 21 (7): 1280–81. https://doi.org/10.1093/bioinformatics/bti141.

Carlsson, Gunnar, and Facundo Mémoli. 2010. "Characterization, Stability and Convergence of Hierarchical Clustering Methods." *The Journal of Machine Learning Research* 11: 1425–70.

Chierchia, Giovanni, and Benjamin Perret. 2019. "Ultrametric Fitting by Gradient Descent." In *Advances in Neural Information Processing Systems*, 3181–92.

Csardi, Gabor, and Tamas Nepusz. 2006. "The Igraph Software Package for Complex Network Research." *InterJournal* Complex Systems: 1695. http://igraph.org.

Fernández, Alberto, and Sergio Gómez. 2008. "Solving Non-Uniqueness in Agglomerative Hierarchical Clustering Using Multidendrograms." *Journal of Classification* 25 (1): 43–65. https://doi.org/10.1007/s00357-008-9004-x.

Hahsler, Michael. 2017. "An Experimental Comparison of Seriation Methods for One-Mode Two-Way Data." *European Journal of Operational Research* 257: 133–43.

Hahsler, Michael, Kurt Hornik, and Christian Buchta. 2008. "Getting Things in Order: An Introduction to the r Package Seriation." *Journal of Statistical Software* 25 (3): 1–34. http://www.jstatsoft.org/v25/i03/.

Hornik, Kurt. 2005. "A CLUE for CLUster Ensembles." *Journal of Statistical Software* 14 (12). https://doi.org/10.18637/jss.v014.i12.

Huettner, Amelie, and Jan-Frederik Mai. 2019. "Simulating Realistic Correlation Matrices for Financial Applications: Correlation Matrices with the Perron–Frobenius Property." *Journal of Statistical Computation and Simulation* 89 (2): 315–36. https://doi.org/10.1080/00949655.2018.1546861.

Jaeger, Markus, Stephan Krügel, Dimitri Marinelli, Jochen Papenbrock, and Peter Schwendner. 2021. "Interpretable Machine Learning for Diversified Portfolio Construction." *The Journal of Financial Data Science* 3 (3).

Jain, Prayut, and Shashi Jain. 2019. "Can Machine Learning-Based Portfolios Outperform Traditional Risk-Based Portfolios? The Need to Account for Covariance Misspecification." *Risks* 7 (3): 74. https://doi.org/10.3390/risks7030074.

Jothimani, D., and A. Bener. 2019. "Risk Parity Models for Portfolio Optimization: A Study of the Toronto Stock Exchange." In *2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)*, 27–32.

Kaufman, Leonard, and Peter J. Rousseeuw, eds. 1990. *Finding Groups in Data.* John Wiley & Sons, Inc. https://doi.org/10.1002/9780470316801.

Kolrep, Martin, Harald Lohre, Erhard Radatz, and Carsten Rother. 2020. "Economic Versus Statistical Clustering in Multi-Asset Multi-Factor Strategies." *Risk & Reward*, 26–32.

Lohre, Harald, Jochen Papenbrock, and Muddit Poonia. 2014. "The Use of Correlation Networks in Parametric Portfolio Policies." *SSRN*. https://doi.org/http://dx.doi.org/10.2139/ssrn.2505732.

Lohre, Harald, Carsten Rother, and Kilian Axel Schäfer. 2020. "Hierarchical Risk Parity: Accounting for Tail Dependencies in Multi-Asset Multi-Factor Allocations." Wiley. https://doi.org/10.1002/9781119751182.ch9.

Lopez de Prado, Marcos. 2016a. "A Robust Estimator of the Efficient Frontier." *SSRN*. https://ssrn.com/abstract=3469961.

———. 2016b. "Building Diversified Portfolios That Outperform Out of Sample." *The Journal of Portfolio Management* 42 (4): 59–69. https://doi.org/10.3905/jpm.2016.42.4.059.

Ma, Xiaofei, and Satya Dhavala. 2018. "Hierarchical Clustering with Prior Knowledge." *arXiv Preprint arXiv:1806.03432*.

Mantegna, R. N. 1999. "Hierarchical Structure in Financial Markets." *The European Physical Journal B-Condensed Matter and Complex Systems* 11: 193–79. https://doi.org/https://doi.org/10.1007/s100510050929.

Marti, G., F. Nielsen, M. Binkowski, and P. Donnat. 2017. "A review of two decades of correlations, hierarchies, networks and clustering in financial markets." *ArXiv e-Prints*, March. http://arxiv.org/abs/1703.00485.

Marti, Gautier. 2019. "CorrGAN: Sampling Realistic Financial Correlation Matrices Using Generative Adversarial Networks." http://arxiv.org/abs/1910.09504.

Menchero, Jose, and Lei Ji. 2019. "Portfolio Optimization with Noisy Covariance Matrices." *Journal Of Investment Management* 17 (1): 200–206.

Molyboga, Marat. 2020. "A Modified Hierarchical Risk Parity Framework for Portfolio Management." *The Journal of Financial Data Science*. https://doi.org/10.3905/jfds.2020.1.038.

Murtagh, Fionn. 2007. "Identifying and Exploiting Ultrametricity." In *Studies in Classification, Data Analysis, and Knowledge Organization*, 263–72. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-70981-7_30.

Müllner, Daniel. 2013. "Fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for r and Python." *Journal of Statistical Software* 53 (9). https://doi.org/10.18637/jss.v053.i09.

Onnela, J.-P., A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto. 2003. "Dynamics of Market Correlations: Taxonomy and Portfolio Analysis." *Phys. Rev. E* 68 (November): 056110. https://doi.org/10.1103/PhysRevE.68.056110.

Pafka, Szilard, and Imre Kondor. 2002. "Noisy Covariance Matrices and Portfolio Optimization." *The European Physical Journal B - Condensed Matter and Complex Systems* 27 (2).

Papenbrock, Jochen. 2011. "Asset Clusters and Asset Networks in Financial Risk Management and Portfolio Optimization." PhD thesis, Karlsruhe. https://doi.org/10.5445/IR/1000025469.

Papenbrock, Jochen, and Peter Schwendner. 2015. "Handling Risk on/Risk Off Dynamics with Correlation Regimes and Correlation Networks." *Financial Markets and Portfolio Management* 29: 2. 125–147.

Papenbrock, Jochen, Peter Schwendner, Markus Jaeger, and Stephan Krügel. 2021. "Matrix Evolutions: Synthetic Correlations and Explainable Machine Learning for Constructing Robust Investment Portfolios." *The Journal of Financial Data Science* 3 (2).

Petrie, WM Flinders. 1899. "Sequences in Prehistoric Remains." *Journal of the Anthropological Institute of Great Britain and Ireland*, 295–301.

Pfitzinger, Johann, and Nico Katzke. 2019. "A constrained hierarchical risk parity algorithm with cluster-based capital allocation." Working Papers 14/2019. Stellenbosch University, Department of Economics. https://ideas.repec.org/p/sza/wpaper/wpapers328.html.

Prado, Marcos López de. 2019. "Estimation of Theory-Implied Correlation Matrices." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3484152.

Raffinot, Thomas. 2017. "Hierarchical Clustering-Based Asset Allocation." *The Journal of Portfolio Management* 44 (2): 89–99. https://doi.org/10.3905/jpm.2018.44.2.089.

———. 2018. "The Hierarchical Equal Risk Contribution Portfolio." *SSRN*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3237540.

Schmid, Friedrich, and Rafael Schmidt. 2007. "Multivariate Conditional Versions of Spearman's Rho and Related Measures of Tail Dependence." *Journal of Multivariate Analysis* 98 (6): 1123–40.

Tola, Vincenzo, Fabrizio Lillo, Mauro Gallegati, and Rosario N. Mantegna. 2008. "Cluster analysis for portfolio optimization." *Journal of Economic Dynamics and Control* 32 (1): 235–58. https://ideas.repec.org/a/eee/dyncon/v32y2008i1p235-258.html.

Tumminello, M, F Lillo, and R. N Mantegna. 2007. "Hierarchically Nested Factor Model from Multivariate Data." *Europhysics Letters (EPL)* 78 (3): 30006. https://doi.org/10.1209/0295-5075/78/30006.

Ward Jr, Joe H. 1963. "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association* 58 (301): 236–44.