Hierarchical risk parity: Accounting for tail dependencies in multi-asset multi-factor allocations*

Harald Lohre[†]
Invesco
EMP, Lancaster University Management School

Carsten Rother[‡]
Invesco
University of Hamburg

Kilian Axel Schäfer[§]
Metzler Asset Management

January 23, 2020

*This manuscript is forthcoming chapter 9 in *Machine Learning and Asset Management*, Elsevier & ISTE Press, edited by Emmanuel Jurczenko. We are grateful to Victoria Beschorner, Emmanuel Jurczenko, Erhard Radatz, Lukasz Szpruch and seminar participants at the 2018 Global Research Meeting of Invesco Quantitative Strategies in Frankfurt. Note that this paper expresses the authors' views which do not necessarily coincide with those of Invesco or Metzler.

[†]Invesco Quantitative Strategies, An der Welle 5, 60322 Frankfurt/Main, Germany; Centre for Financial Econometrics, Asset Markets and Macroeconomic Policy (EMP), Lancaster University Management School, Bailrigg, Lancaster LA1 4YX, United Kingdom; harald.lohre@invesco.com

[‡]Contact Author: Invesco Quantitative Strategies, An der Welle 5, 60322 Frankfurt/Main, Germany; University of Hamburg, Moorweidenstraße 18, 20148 Hamburg, Germany; carsten.rother@invesco.com

[§]Metzler Asset Management, Untermainanlage 1, 60329 Frankfurt/Main, Germany; kilian.axel.schaefer@metzler.com

Hierarchical risk parity: Accounting for tail dependencies in multi-asset multi-factor allocations

Abstract

We investigate portfolio diversification strategies based on hierarchical clustering. These hierarchical risk parity strategies use graph theory and unsupervised machine learning to build diversified portfolios by acknowledging the hierarchical structure of the investment universe. In this chapter, we consider two dissimilarity measures for clustering a multi-asset multi-factor universe. While the Pearson correlation coefficient is a popular choice, we are especially interested in a measure based on the lower tail dependence coefficient. Such innovation is expected to achieve better tail risk management in the context of allocating to skewed style factor strategies. Indeed, the corresponding hierarchical risk parity strategies seem to have been navigating the associated downside risk better, yet come at the cost of high turnover. A comparison based on block-bootstrapping evidences alternative risk parity strategies along economic factors to be on par in terms of downside risk with those based on statistical clusters.

Keywords: Multi-asset Multi-factor Investing, Diversification, Hierarchical Risk Parity, Tail dependence

JEL Classification: G11, D81

In an attempt to construct more efficient and better risk-managed portfolios, investors have recently turned to diversifying their portfolios through factors rather than traditional asset classes. We investigate meaningful ways of generating a coherent multi-asset multi-factor allocation that is able to harvest the associated asset and factor premia in a balanced fashion. Standard portfolio theory would suggest to aim for an optimal risk and return trade-off by resorting to the seminal mean-variance paradigm of Markowitz (1952). Yet, given the notorious sensitivity of such portfolio optimization with regard to expected return inputs, we may rather disregard forecasting returns and focus on estimating risk. As a result, researchers developed various risk-based allocation strategies in pursuit of portfolio diversification.

An innovative risk-based strategy for managing diversification was introduced by Meucci (2009). Conducting a principal component analysis (PCA), he aims to identify the main risk drivers in a given set of assets. The orthogonal eigenvectors resulting from the PCA can be viewed as principal portfolios representing uncorrelated risk sources, and a portfolio is considered well-diversified if the overall risk is distributed equally across these uncorrelated principal portfolios. Given the statistical nature of PCA, Meucci, Santangelo, and Deguest (2015) suggest to resort to a minimum-torsion transformation instead. The resulting uncorrelated risk sources are economically more meaningful. Along these lines, the literature has advanced diversified risk parity strategies designed to maximize diversification benefits across asset classes and style factors, see Lohre, Opfer, and Ország (2014) and Bernardi, Leippold, and Lohre (2018).

While such an approach and its outcome are dependent on choosing and designing an appropriate risk model, the recent literature has presented risk parity allocation paradigms guided by hierarchical clustering techniques—prompting Lopez de Prado (2016) to label the technique hierarchical risk parity (HRP). Given a set of asset class and style factor returns, the corresponding algorithm would cluster these according to some distance metric and then allocate equal risk budgets along these clusters. Such clusters might be deemed more natural building blocks than the aggregate risk factors in that they automatically pick up the dependence structure and form meaningful ingredients to aid portfolio diversification.

The contribution of this chapter is to thoroughly examine the use and merits of hierarchical clustering techniques in the context of multi-asset multi-factor investing. In particular, it will contrast these techniques to several competing risk-based allocation paradigms, such as $1/N$, minimum-variance, standard risk parity, and diversified risk parity. A major innovation is to investigate HRP strategies based on tail dependence clustering as opposed to standard correlation-based clustering. Such an approach might be particularly relevant given the elevated tail risk of some style factors. Hierarchical risk parity strategies generally build on two steps: first, hierarchical clustering algorithms uncover a hierarchical structure of the considered investment universe, resulting in a tree-based representation. Second, the portfolio weights result from applying an allocation strategy along the hierarchical structure and thus the overall portfolio is expected to exhibit a meaningful degree of diversification.

As estimates of covariance or correlation matrices are subject to estimation errors, correlation-based clusters and networks are meaningful for constructing diversified portfolios, because these can have a filtering effect, resulting in more reliable outcomes, see Tumminello, Lillo, and Mantegna (2010). In this vein, Lopez de Prado (2016) argues that the correlation matrix is too complex to be fully analyzed and lacks a hierarchical structure. Instead of analyzing the full covariance matrix, he suggests to consider the corresponding minimum spanning tree (MST) with N nodes (one node for each asset) and only $N-1$ edges, i.e. focusing on the most relevant correlations. Deriving the MST requires the definition of a distance measure, which is often referred to as a dissimilarity measure. The MST is naturally linked to the hierarchical clustering algorithm, called single linkage. In a direct way, the MST gives the hierarchical organization of the investigated assets and the optimal portfolio weights can be derived by applying an allocation scheme to the hierarchical structure.

Given a multi-asset multi-factor investment universe, it has to be kept in mind that the return distributions of the underlying factor strategies can be highly skewed. A well-known investment style is the FX carry trade that stipulates buying currencies with the highest short-term interest rates and selling those with the lowest short-term interest rates. While the return pattern of the FX carry trade is benign in calm markets, it happens to be prone to sudden drawdowns in stress episodes, when FX investors shy away from more risky currencies. Obviously, this is a prime example as to why tail risk is an important consideration in the context of constructing style factor allocations. Therefore, this chapter considers portfolio construction techniques that specifically incorporate the notion of tail risk management. While the above risk parity allocation paradigm is ultimately centered around maintaining equal-volatility risk budgets, we may wonder as to whether alternative approaches might be more appropriate to navigate the non-normality of factor returns. In that regard, Boudt, Carl, and Peterson (2013) put forward equalizing risk contributions to the portfolio's CVaR.[1] This approach is thus avoiding risk concentrations in the portfolio's tail risk and could simply be thought of as tail risk parity. Yet, the underlying portfolio optimization is more cumbersome and less robust than the traditional risk parity algorithm. In that sense, we will seek to apply the above robust hierarchical clustering based on tail dependency measures to explicitly account for tail risk while not sacrificing robustness of portfolio weights.

A variety of other allocation strategies based on hierarchical clustering were developed using different linkage methods, allocation schemes and (dis-)similarity measures; see, for instance, Raffinot

[1] Recently, Jurczenko and Teiletche (2019) have proposed an expected shortfall risk budgeting framework that accounts for risks beyond volatility, including asymmetry and tail risk.

(2017). These measures usually build on the correlation coefficient. To acknowledge tail risk, other scholars proposed measures focusing on the association among assets during times of large losses. One possible choice is the lower tail dependence coefficient; see, for instance, De Luca and Zuccolotto (2011) and Durante, Pappadà, and Torelli (2015). Building upon hierarchical risk parity of Lopez de Prado (2016), we propose and investigate innovative variations of this strategy using the lower tail dependence coefficient.

The chapter is organized as follows: Section 1 illustrates hierarchical clustering in the context of the multi-asset multi-factor universe. Further, we introduce hierarchical risk parity strategies based on the Pearson correlation coefficient. Section 2 introduces hierarchical clustering based on the lower tail dependence coefficient. Section 3 provides an overview of traditional risk-based allocation strategies and outlines a framework to measure and manage portfolio diversification. Section 4 examines the performance of the introduced HRP strategies relative to the traditional alternatives. Section 5 concludes.

1 Hierarchical risk parity strategies

Investigating hierarchical risk parity strategies in the context of multi-asset multi-factor investing, the respective market and style factor universe is discussed. Afterwards, we illustrate the concept of hierarchical risk parity by conveying the inherent hierarchical structure.

1.1 The multi-asset multi-factor universe

We consider multi-asset multi-factor data that combines the traditional asset classes equity, bond, commodity and credit as well as different style factors. The time series are available for the period from March 3, 2003, to December 29, 2017. The global equity and bond markets are represented by equity index futures for S&P 500, Nikkei 225, FTSE 100, EuroSTOXX 50, MSCI Emerging Markets and bond index futures for US 10Y Treasuries, Bund, Japanese Government Bonds (JGB) 10Y and U.K. gilts. The credit risk premium is captured by the Bloomberg Barclays US Corporate Investment Grade (Credit IG) and High Yield (Credit HY) Indices, both are duration-hedged to synthesize pure credit risk. Further, gold, oil, copper and agriculture indices are chosen to cover the commodity market.

Style factors systematically follow specific investment styles. In total, we consider the five investment styles carry, value, momentum, quality and volatility carry within the asset classes equity, (interest) rates, commodity and foreign exchange. Carry is based on the idea that high-yield assets tend to outperform low-yield assets, while momentum investors assume that recent winning assets outperform recent losing assets. Quality (or defensive) investing builds on the observation that high-quality assets tend to have higher risk-adjusted returns than low-quality assets. Value investing is based on the idea that cheap assets (according to a given valuation metric) tend to outperform expensive assets. Finally, the volatility carry style is based on the assumption that long-volatility assets will decline in value due to an upward sloping forward volatility curve. We source the underlying return time series from Goldman Sachs (GS) and Invesco Quantitative Strategies (IQS). The exact factor definitions can be found in Appendix A, as provided by GS and IQS.

For measuring diversification, an appropriate factor model along suitable economic factors can be established. To benchmark the statistical clusters vis-à-vis economic factors, we will include the market factors equity + credit, duration and commodity. Further, taking a pure style factor investing perspective, we build aggregate style factors across asset classes, i.e. the aggregate momentum style factor would be based on equity momentum, FX momentum, rates momentum and commodity momentum. In the same vein, we construct aggregate carry, value, quality and volatility carry factors. The aggregated market and style factors each apply a weighting scheme that ensures equal risk contributions of the underlying asset classes or style factors.

1.2 The hierarchical multi-asset multi-factor structure

Many portfolio optimization methods like the Markowitz mean-variance approach are sensitive to changes in input variables, and small estimation errors can lead to vast differences in optimal portfolio allocations. Lopez de Prado (2016) asserts that correlation and covariance matrices are simply too complex to be fully analyzed and disregard the hierarchical structure of interactions among assets. The full variance-covariance matrix can be visualized by a complete graph with N nodes and $\frac{1}{2}N(N-1)$ edges, see Figure 1 (left chart). To reduce complexity, we would want to focus on relevant correlations by dropping unnecessary edges. In this regard, a well-known approach from graph theory is the so-called minimum spanning tree (MST). The MST is a subset of $N-1$ edges of the original graph, connecting all vertices without any cycles and with the minimum total edge weight. An algorithm for obtaining the MST was introduced by Prim (1957). Before applying this algorithm, we have to define a distance measure, which is often based on the correlation coefficient. We will refer to this measure as the dissimilarity measure, since it aims to measure the dissimilarity of the assets (and factors). A precise definition will follow in the next sub-section. Applying the dissimilarity measure to the correlation matrix leads to the so-called dissimilarity matrix and allows us to derive the MST; see Figure 1 (right chart). As this is based on correlations, it is often referred to as correlation network.

We successfully reduced the complete graph with $\frac{1}{2}N(N-1)$ edges to a connected graph with $N-1$ edges, which focuses on essential information contained in the correlation matrix and introduces

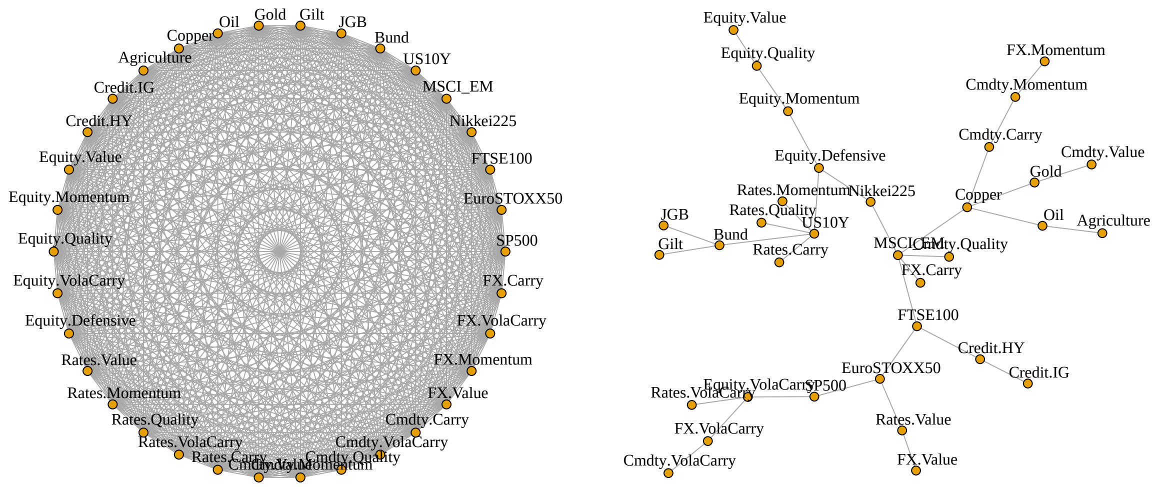

Figure 1: Complete graph (left) versus minimum spanning tree (right)

a hierarchical structure. Looking at the branches, one identifies a commodity cluster (top right), an equity style cluster (top left) and a volatility style cluster (bottom left).

Graph theory is linked to unsupervised machine learning. In particular, the MST is naturally related to the hierarchical clustering algorithm, called single linkage. In a direct way, the MST conveys the hierarchical organization of the investigated assets and style factors; see Mantegna (1999) for early applications in equity universes. This results in a tree structure as represented by the dendrogram in Figure 2. Moving up the tree, objects that are similar to each other are combined into branches, i.e. the higher the height of the fusion, the less similar the objects are.

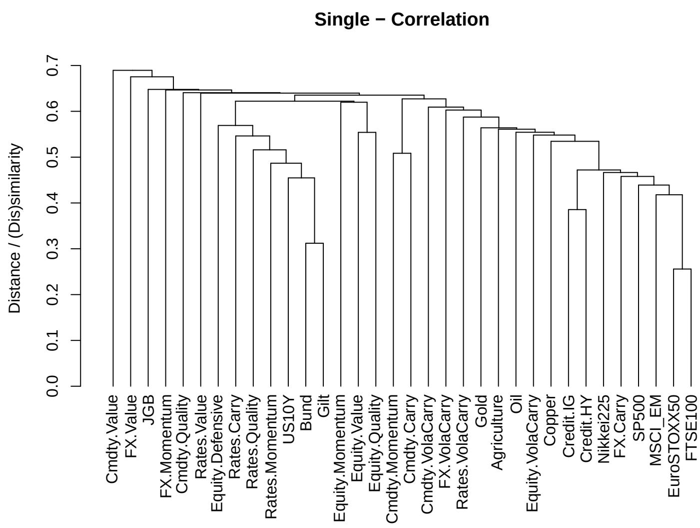

Figure 2: Dendrogram (single linkage) associated with MST from Figure 1

Another way of visualizing the embedded hierarchical structure is comparing the "unorganized" original correlation matrix to the correlation matrix resulting from reordering columns and rows according to the derived dendrogram; see Figure 3. This reordering results in a block-diagonal organization of the covariance matrix such that similar assets and style factors are placed together, whereas dissimilar assets (or factors) are placed further apart. This feature seems certainly helpful for discerning meaningful building blocks to diversify portfolios.

Using the ensuing order of the assets, Lopez de Prado (2016) suggests an allocation technique based on recursive bisection and inverse variance allocation. Starting with one set of ordered assets and unit weights, he bisects the sets in each step and re-scales the weights of the assets in the resulting subsets with a factor α and $(1-\alpha)$, respectively. The factor α is obtained by calculating the inverse variance allocation of the two subsets, where the variance of the subsets is determined by also using an inverse variance allocation within the subset. The procedure is summarized in Algorithm 1.

Algorithm 1 Recursive Bisection

1. Initialization
   (a) Set the list of items: $L = \{L_0\}$, with $L_0 = \{n\}_{n=1,...,N}$ according to the order obtained by the dendrogram
   (b) Assign a unit weight to all assets: $\omega_n = 1,\ \forall n = 1, ..., N$
2. If $|L_i| = 1,\ \forall L_i \in L$, then stop
3. For each $L_i \in L$ such that $|L_i| > 1$:
   (a) Bisect L_i into two subsets, $L_i^{(1)} \cup L_i^{(2)} = L_i$, where $|L_i^{(1)}| = \text{round}\left(\frac{1}{2}|L_i|\right)$, and the order is preserved
   (b) Define the variance of $L_i^{(j)}$, $j = 1, 2$ as $\tilde{V}_i^{(j)} = (\tilde{\omega}_i^{(j)})^T V_i^{(j)} \tilde{\omega}_i^{(j)}$, where $V_i^{(j)}$ is the covariance matrix between the constituents of the $L_i^{(j)}$ bisection and $\tilde{\omega}_i^{(j)}$ are the inverse variance weights
   (c) Compute the split factor: $\alpha_i = 1 - \frac{\tilde{V}_i^{(1)}}{\tilde{V}_i^{(1)} + \tilde{V}_i^{(2)}}$
   (d) Re-scale allocations ω_n by a factor of α_i, $\forall n \in L_i^{(1)}$
   (e) Re-scale allocations ω_n by a factor of $(1 - \alpha_i)$, $\forall n \in L_i^{(1)}$
4. Loop to 2.

The introduced methodology of allocating according to hierarchical clustering is very flexible and one can conceive many variations thereof. For instance, various dissimilarity measures and hierarchical clustering methods can be used. Further, different functions for obtaining $\tilde{\omega}$ and α in steps $3.\,b$) and $3.\,c$) of Algorithm 1 can be employed. Yet, by using recursive bisection Lopez de Prado ignores the structure of the clusters at the different levels of the dendrogram and rather focuses on the order obtained at the bottom level. As this introduces a certain arbitrariness to the allocation algorithm, we will advance Algorithm 1 in a way that would split the allocations using the clusters obtained from the dendrogram.

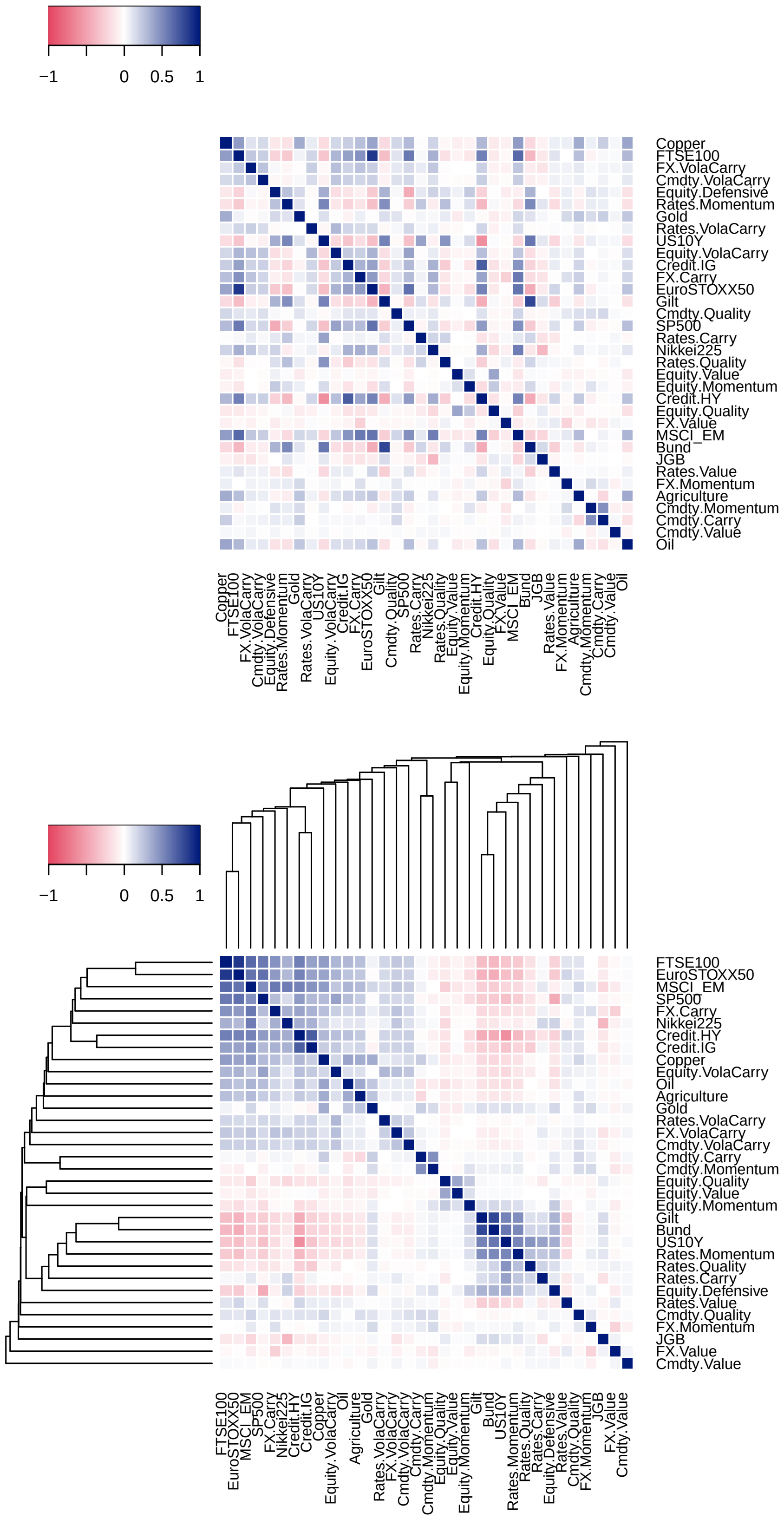

Figure 3: “Random” correlation matrix versus quasi-diagonalized correlation matrix

1.3 Hierarchical clustering

Next, we will formally introduce hierarchical clustering by defining the dissimilarity measure and different linkage-methods. Further, we will state an algorithm for obtaining the portfolio weights based on the dendrogram. We will assume that X is a $T \times N$ data matrix containing T observed returns for each of the N assets. X_j will denote the return times series of asset j.

Cluster analysis aims to group objects in clusters such that objects in a given cluster tend to be similar and objects in different clusters tend to be dissimilar. In particular, hierarchical clustering algorithms find recursively nested clusters, resulting in a tree-based representation called dendrogram. Agglomerative clustering, also known as AGNES (agglomerative nesting), is the most commonly used hierarchical clustering method. Starting with each object being a single-element cluster, the algorithm works in a bottom-up manner, i.e. it sequentially merges the two most similar clusters until a single cluster contains all objects.

The notion of (dis)similarity between the considered objects is key for cluster analysis. Dissimilarity-based clustering algorithms, such as AGNES, require replacing the $T \times N$ data matrix X by an $N \times N$ dissimilarity matrix D. D obtains by applying a dissimilarity measure to the data matrix X.

Definition 1.1. ***(Dissimilarity measure)*** *Let X_i $i = 1, ..., N$ be the columns of a data matrix X, each containing T observations. A function $d : B \rightarrow [0, \infty)$ is called dissimilarity measure if*

$$d(X_i, X_j) \geq 0, \quad d(X_i, X_j) = d(X_j, X_i) \quad \textit{and} \quad d(X_i, X_i) = 0 \quad \forall i, j = 1, ...N, \tag{1}$$

where $B = \{(X_i, X_j)|i, j = 1, ..., N\}$.

Note that the dissimilarity matrix D defines a metric space, if function d additionally satisfies the triangle inequality. There are numerous ways of defining the dissimilarity measure including Euclidean and Manhattan distances, but for the moment we will consider $d : B \rightarrow [0, 1]$ as defined by Lopez de Prado (2016):

$$d_{i,j} = d(X_i, X_j) = \sqrt{0.5(1 - \rho_{i,j})}, \tag{2}$$

where $\rho_{i,j} = \rho_{i,j}(X_i, X_j)$ is the Pearson correlation coefficient. We can verify that (2) is a dissimilarity measure, see, for instance, Lopez de Prado (2016). For perfectly correlated assets ($\rho_{i,j} = 1$), we have $d = 0$. For perfectly negatively correlated assets ($\rho_{i,j} = -1$), we have $d = 1$.

D contains the information about the dissimilarity among all single assets and factors. Still, we have to define how to use this information for measuring the (dis)similarity among clusters containing more than one element. This is done by the so-called linkage criterion. There are various criterions, but the most common ones are single linkage, complete linkage, average linkage and Ward's method, see, for instance, Raffinot (2017).

In the case of *single linkage* the distance between two clusters is defined as the minimal distance of any two elements in the clusters:

$$d_{C_i,C_j} = \min_{x \in C_i, y \in C_j} d(x,y), \tag{3}$$

where C_i and C_j denote two clusters and x and y represent elements within the clusters. By construction, this linkage criterion is sensitive to outliers and can lead to long and straggly clusters, which is a well-known problem called chaining. *Complete linkage* in contrast is defined as the maximum distance between any two elements of the clusters:

$$d_{C_i,C_j} = \max_{x \in C_i, y \in C_j} d(x,y) \tag{4}$$

The resulting clusters tend to be compact with a similar size. However, the strategy is also sensitive to outliers. Using *average linkage*, one computes the average of the distance of any two elements in the examined clusters:

$$d_{C_i,C_j} = \frac{1}{|C_i|}\frac{1}{|C_j|}\sum_{x \in C_i}\sum_{y \in C_j} d(x,y), \tag{5}$$

where $|C_i|$ is the cardinality of cluster i. *Ward's method* minimizes the total within-cluster variance:

$$d_{C_i,C_j} = \frac{|C_i||C_j|}{|C_i| + |C_j|}|\mu_i - \mu_j|^2, \tag{6}$$

where μ_i is the center of cluster i. This method aims to find compact clusters and is considered to be less sensitive to outliers.

Deriving the dendrogram, one starts with each asset (or factor) being in a single-element cluster. Then, based on the information provided by the dissimilarity matrix and the chosen linkage criterion, the two most similar clusters are sequentially merged. The dendrograms resulting from the four linkage methods can be observed in Figure 4: Ward's method results in compact clusters of similar size, making it a popular choice among researchers. Conversely, single linkage suffers from chaining, and the other methods are sensitive to outliers.

The dendrogram shows the merging of clusters in each step of the algorithm. Moving up the dendrogram, similar objects/clusters are combined into branches. On the axis, we provide the value at which the merge occurs, representing the (dis)similarity between two objects/clusters. However, we can only draw conclusions about the proximity of two objects if they are merged. The dendrogram does not allow for inferences about the proximity of two "random" objects/clusters.

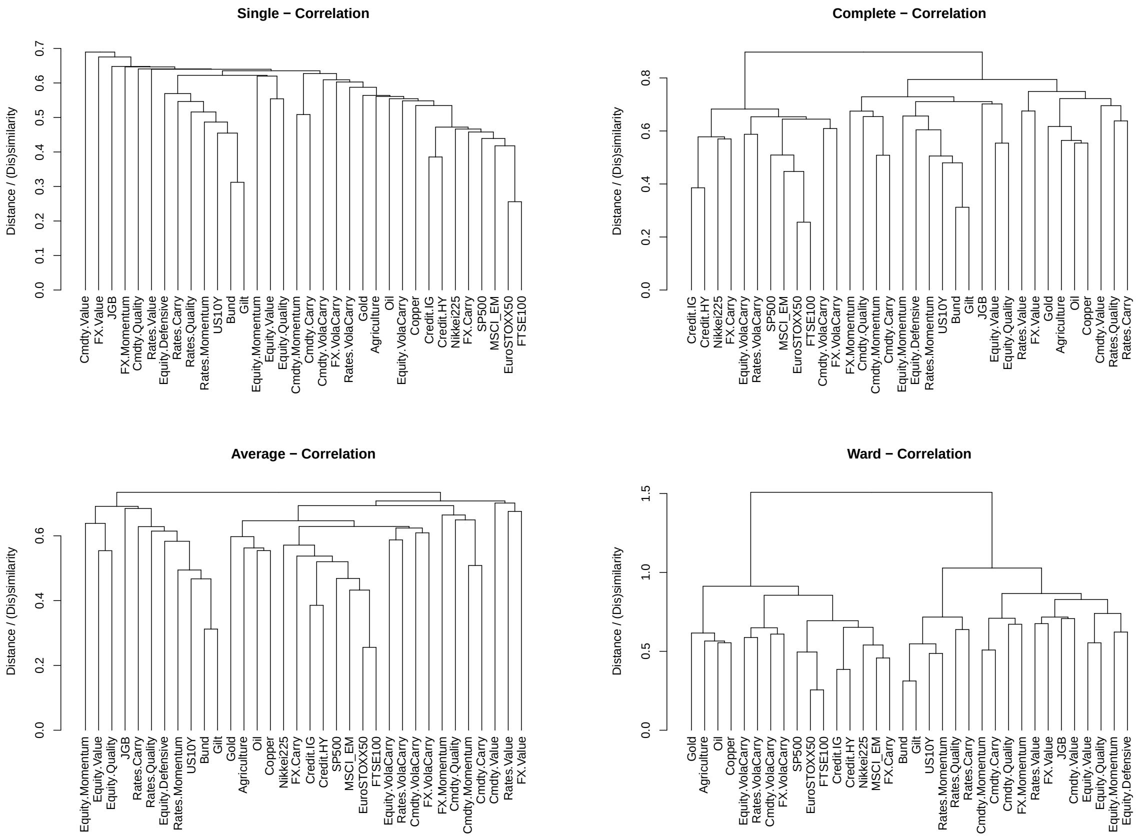

Figure 4: Comparison of linkage methods: single linkage (top left), complete linkage (top right), average linkage (bottom left) and Ward's method (bottom right)

In the case of a large-investment universe, the dendrogram might want to be considered only up to a certain level instead of taking into account the whole hierarchical structure. While this reduction leads to a loss of information, finding the weight allocation is faster. Cutting the dendrogram will partition the assets and style factors into clusters. There are different ways to determine an optimal number of clusters. We could simply choose a plausible number by looking at the dendrogram or apply a statistical criterion for determining the "optimal" number of clusters. An example is given in Figure 5, where the number of clusters was deliberately chosen to be seven. Popular statistical methods are the gap statistic [Tibshirani, Walther, and Hastie (2001)] and the silhouette index [Rousseeuw (1987)].

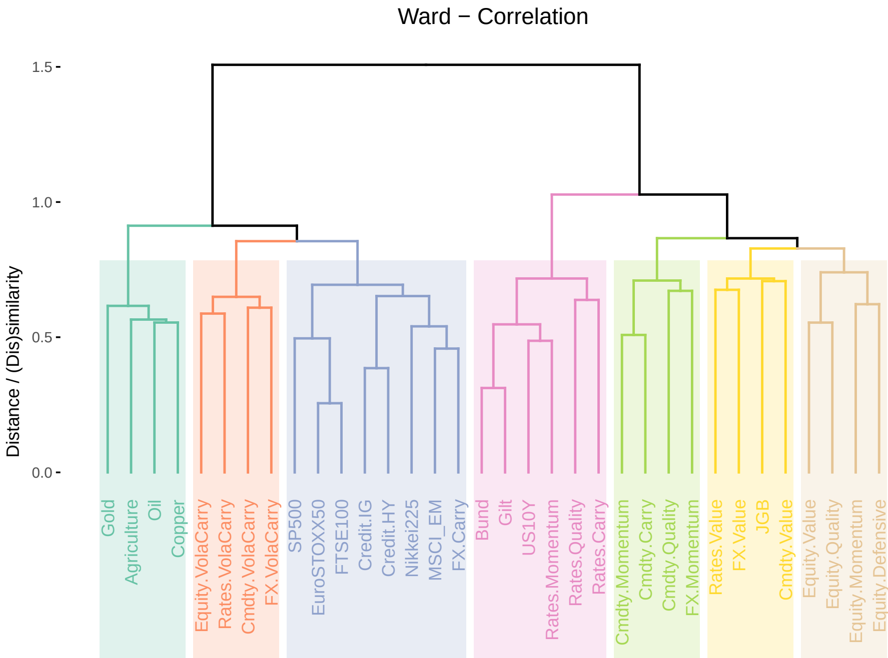

Figure 5: Dendrogram based on Ward's method; cut such that seven clusters obtain

1.4 Portfolio allocation based on hierarchical clustering

Having determined the dendrogram, we have to decide how to allocate one's capital. Instead of using an algorithm based on recursive bisection, we rather want to invest along the nodes of the dendrogram to acknowledge the hierarchical information. Further, we have to choose an allocation technique within and across clusters. Lopez de Prado uses the inverse variance strategy in both cases, yet there are various alternatives available. For instance, Papenbrock (2011) and Raffinot (2017) suggest a weighting scheme that allocates capital equally across cluster hierarchy and within clusters. We will use different combinations of risk parity strategies, including inverse volatility and equal risk contribution. The general procedure is described in Algorithm 2 where we have to choose a methodology for the within- and across-cluster allocation. The algorithm starts at the top of the dendrogram and assigns weights by going from node to node. Note that step 4 is only to be executed if an optimal number of clusters is used, i.e. not all remaining clusters are singleton clusters.

Algorithm 2 Clustering-based weight allocation

1. Perform hierarchical clustering and generate dendrogram
2. Assign all assets a unit weight $\omega_i = 1 \; \forall i = 1, ..., N$
3. For each dendrogram node (beginning from the top):
   (a) Determine the members of clusters C_1 and C_2 belonging to the two sub-branches of the according dendrogram node
   (b) Calculate the within cluster allocations $\tilde{\omega}_1$ and $\tilde{\omega}_2$ for C_1 and C_2 according to the chosen methodology
   (c) Based on the within cluster allocations $\tilde{\omega}_1$ and $\tilde{\omega}_2$ calculate the across cluster allocation α (splitting factor) for C_1 and $(1 - \alpha)$ for C_2 according to the chosen methodology
   (d) For each asset in C_1 re-scale allocation ω by factor α
   (e) For each asset in C_2 re-scale allocation ω by factor $(1 - \alpha)$
4. For each cluster containing more than one element:
   (a) Determine the members of the cluster
   (b) Calculate the within cluster allocation
   (c) For each asset in the cluster re-scale ω by the within cluster allocation
5. End

2 Tail dependency and hierarchical clustering

So far, we have investigated hierarchical clustering based on the correlation matrix. However, correlations can increase significantly during financial crises because of contagion effects. Financial contagion can be triggered by unexpected extreme events and spread quickly. The best-known example is the Great Depression, which was triggered by the U.S. stock market crash in 1929 and affected economies around the globe. Hence, diversification by correlation clusters may fail to work when it is needed most.

Alternatively, one might group assets according to their co-movement when they experience large losses. Such an approach was proposed by De Luca and Zuccolotto (2011), who consider lower tail dependence coefficients (LTDC) to capture the pairwise tail behavior of the underlying assets. While the correlation coefficient describes the association of the entire distribution, the LTDC quantifies the association in extreme negative events. Presumably, lower tail dependence is more directly linked to the ability to diversify in crises, potentially improving the tail risk management of a given strategy.

2.1 Tail dependence coefficients

In the following, we formally introduce the lower and upper tail dependence coefficients and define a dissimilarity measure based on the LTDC. The main idea is to cluster assets that are characterized by a high probability of experiencing extremely negative events contemporaneously. The tail dependence coefficients are defined by a conditional probability: the probability that a random variable X takes an extreme value, given that an extreme value has occurred for random variable Y.

Definition 2.1. *(Tail dependence coefficients) Let X and Y be two random variables and let $F_X(x) = \mathbb{P}(X \leq x)$ and $F_Y(y) = \mathbb{P}(Y \leq y)$ be their distribution functions. The lower tail dependence coefficient (LTDC) is defined as*

$$\lambda_L = \lim_{t \searrow 0} \mathbb{P}(X \leq F_X^{-1}(t) | Y \leq F_Y^{-1}(t)) \tag{7}$$

and the upper tail dependence coefficient (UTDC) is

$$\lambda_U = \lim_{t \nearrow 1} \mathbb{P}(X > F_X^{-1}(t) | Y > F_Y^{-1}(t)). \tag{8}$$

We call X and Y asymptotically dependent (independent) in the lower tail if $0 < \lambda_L \leq 1$ ($\lambda_L = 0$) and accordingly for λ_U. Quantifying this probability solely depends on the assumed bivariate distribution, i.e. the existence of tail dependence is a property of the bivariate distribution, see De Luca and Zuccolotto (2011).

In applications, data is needed to estimate the tail dependence coefficients based on a distributional assumption. The latter being rather difficult for financial data, copula functions are widely used instead. Copulas allow to describe the joint distribution function in terms of univariate marginal distribution functions. As a result, copulas allow to effectively estimate the tail dependence in a simple and flexible manner. Assuming we can estimate the LTDC, we define the dissimilarity measure as suggested by De Luca and Zuccolotto (2011).

Proposition 2.1. *Let X and Y be two random variables and λ_L their associated lower tail dependence coefficient, then the following function, $d : \mathbb{R} \times \mathbb{R} \to [0, \infty)$, is a dissimilarity measure:*

$$d(X, Y) = -\log(\lambda_L). \tag{9}$$

2.2 Estimating tail dependence coefficients

Copulas are popular in financial applications since they allow us to separate the analysis of a joint distribution function into two parts. First, the marginal distribution functions are analyzed independently for each random variable. Second, we choose a copula model to describe the dependence structure as a function of the margins, or we use an empirical version of the copula. Hence, a copula links the marginal distributions and the joint distribution and implicitly defines a bivariate distribution function.

One mainly distinguishes between parametric and non-parametric approaches for estimating copula-based tail coefficients. De Luca and Zuccolotto (2011), who first introduced a clustering approach based on the LTDC, used a parametric approach based on Archimedean copulas. Having selected a model and having estimated the according parameters, the LTDC can be easily computed. To avoid specifying a parametric model assumption on the pairwise dependence structure of the involved time series, we will briefly discuss two non-parametric approaches.

The procedure of Durante, Pappadà, and Torelli (2015) is based on an extreme value theory approach that requires defining copulas, survival copulas and extreme value copulas. Moreover, we can establish a connection between these different concepts and the lower and upper tail dependence coefficients. Yet, in unreported results, we have found the dendrograms arising from this estimator to be rather unstable. As portfolio allocations along such dendrograms will suffer from high turnover, we are resorting to the more stable non-parametric estimator of Schmid and Schmidt (2007). The authors propose a multi-dimensional generalization of the LTDC based on a conditional version of Spearman's rho, which weights the different parts of the multidimensional copula. We therefore refer to this estimator as the CSR-estimator (Conditional Spearman's Rho). For our purpose of estimating the pairwise LTDC, we are considering the two-dimensional version. See Appendix B for the definition of the CSR-estimator and Schmid and Schmidt (2007) for its derivation.

Figure 6 shows a dendrogram based on (9) where the CSR-Estimator is used for estimating the LTDCs. In contrast to Figure 5, one can observe that the dissimilarity measures can lead to fairly different dendrograms. Note that we choose the optimal number of clusters to be five; other choices are conceivable. Intuitively, using the LTDC would see a joint cluster of commodity market and commodity style factors. Relative to the correlation-based dendrogram in Figure 5 we find the equity-credit cluster enlarged, including more style factors. Interestingly, equity style factors form one cluster, except for Defensive Equity that clusters with the Japanese stock index.

In fact, there are various ways to define a dissimilarity measure based on the LTDC. Definition (9) is a popular choice for the following reasons. It ranges from 0 to infinity, since $\lambda_L \in [0, 1]$. Moreover, (9) is small when tail dependence is high and monotonically increases as tail dependence decreases. Hence, assets with high tail dependence appear to be "close" to each other, whereas assets with low tail dependence appear to be "more distant" from each other.

3 Risk-based allocation strategies

This section introduces risk-based allocation techniques. First, we focus on "traditional" allocation strategies, used to benchmark the hierarchical risk parity strategies. Second, we discuss Meucci's approach to managing diversification, that serves to construct a diversified risk parity strategy based on economic factors. The latter framework represents a useful benchmark to contrast hierarchical risk

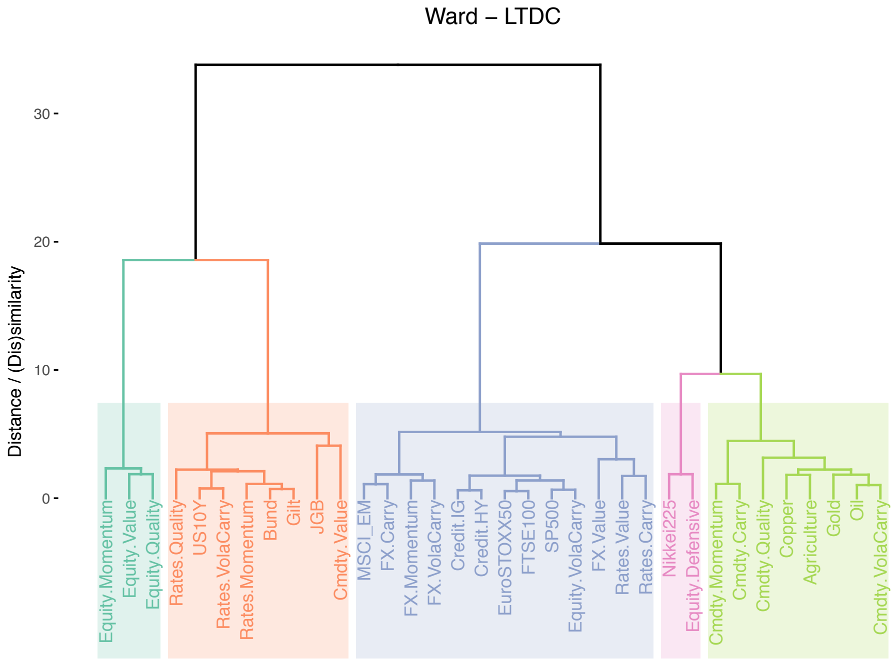

Figure 6: Dendrogram (Ward method) - LTDC-based

parity strategies in terms of diversification.

3.1 Classic risk-based allocation techniques

The most simple allocation strategy is equal weighting ($1/N$). It gives the same importance (in terms of dollars invested) to each asset. For N assets, we obtain weights

$$\omega_i = \frac{1}{N} \quad \text{for } i = 1, ..., N. \tag{10}$$

Another popular risk-based allocation technique is the minimum-variance portfolio (MVP). It is a special case of the Markowitz mean-variance approach, determining the leftmost point on the efficient frontier. Its weights are determined by solving the quadratic optimization problem

$$\arg\min_{\omega} \frac{1}{2} \omega^T \Sigma \omega, \tag{11}$$

where Σ is the covariance matrix. We impose full investment and long-only investment constraints. While minimum-variance portfolio weights are easy to compute, they often concentrate in low-volatility assets. Conversely, $1/N$ has a completely balanced weights allocation, but might suffer from a concentrated risk allocation, if the individual assets' risk differ significantly. Therefore, we may consider diversifying risk by allocating such that all assets contribute equally to overall portfolio risk. Strategies following this objective are called risk parity strategies. A naive approach to risk parity would first weight each asset inversely proportional to its volatility σ_i and, second, rescale weights (letting them add to 1):

$$\omega_i = \frac{1/\sigma_i}{\sum_{j=1}^{N} 1/\sigma_j} \quad \text{for } i = 1, ..., N. \tag{12}$$

This inverse volatility strategy might be considered a naive risk parity strategy, because it is agnostic with respect to asset correlations. Therefore, we next discuss a risk parity strategy that assigns weights such that all assets truly contribute equally to total portfolio risk. A key metric is the marginal risk contribution, which measures the change in portfolio risk σ_p caused by an infinitesimal change in the weight of asset i

$$MRC_i = \partial_{\omega_i}\sigma_p = \frac{(\Sigma\omega)_j}{\sqrt{\omega^T\Sigma\omega}}, \tag{13}$$

where Σ is the covariance matrix. The total risk contribution is then given by $TRC_i = w_i MRC_i$, and we can decompose the portfolio risk as follows

$$\sigma_p = \sum_{i=1}^{N} TRC_i. \tag{14}$$

Hence, a risk parity strategy with equal risk contributions must satisfy $TRC_i = TRC_j$ for all i, j. Since MRC_i and $(\Sigma\omega)_i$ are proportional, we might alternatively require $\omega_i(\Sigma\omega)_i = \omega_j(\Sigma\omega)_j$ for all i, j. However, there is no closed-form solution for the general case. We follow Maillard, Roncalli, and Teiletche (2010) and solve a numerical optimization problem in order to obtain the portfolio weights:

$$\arg\min_{\omega} \sum_{i=1}^{N}\sum_{j=1}^{N} \left(\omega_i(\Sigma\omega)_i - \omega_j(\Sigma\omega)_j\right)^2. \tag{15}$$

3.2 Diversified risk parity

Striving for a well-diversified portfolio, Meucci (2009) constructs uncorrelated risk sources embedded in the underlying portfolio assets. A well-diversified portfolio would follow a risk parity strategy applied to these uncorrelated risk sources. To measure the average number of uncorrelated bets in a portfolio, he introduces the concept of effective number of bets. For constructing uncorrelated risk sources Meucci (2009) suggests using principal component analysis (PCA), while follow-up research of Meucci, Santangelo, and Deguest (2015) advocates a minimum-torsion transformation to derive the linear orthogonal transformation that is closest to the original assets (or a pre-specified factor model). In the following, we will consider this framework in the context of a factor model (rather than individual assets).

3.2.1 Effective number of bets

We are considering k correlated factors with return vector F. Thus, the portfolio return is given by

$$R_p = \sum_{i=1}^{k} b_i F_i, \tag{16}$$

where b_i represents the portfolio weight of the i-th factor. Assuming that the portfolio can be expressed as a combination of uncorrelated factors $\tilde{F}$

$$R_p = \sum_{i=1}^{k} \tilde{b}_i \tilde{F}_i, \tag{17}$$

we can compute the relative contribution of each uncorrelated factor to total risk:

$$p_i = \frac{Var(\tilde{b}_i \tilde{F}_i)}{Var(R_p)} = \frac{\tilde{b}_i^2 Var(\tilde{F}_i)}{\sum_{i=1}^{k} \tilde{b}_i^2 Var(\tilde{F}_i)}, \quad i = 1, ..., k. \tag{18}$$

Notice that the p_i sum to 1 and are non-negative. Hence, Meucci calls (18) the diversification distribution. He conceives a portfolio to be well-diversified if the diversification distribution is uniform, and uniformity is maximal for a risk parity strategy along the uncorrelated bets. To assess the degree of uniformity Meucci defines the effective number of bets as the exponential of the negative entropy

$$N_{Ent} = \exp\left(-\sum_{i=1}^{k} p_i \ln(p_i)\right). \tag{19}$$

To foster intuition, consider two extreme cases. A portfolio concentrated in a single factor leads to $p_i = 1$ for one i and $p_j = 0$ for $i \neq j$, resulting in N_{Ent}=1, a single bet. Vice versa, a well-diversified portfolio with $p_i = \frac{1}{N}$ for all i gives $N_{Ent} = k$, the maximum number of uncorrelated bets.

Meucci, Santangelo, and Deguest (2015) give an explicit expression for the effective number of bets, letting t be a linear transformation such that

$$\tilde{F} = tF, \quad \tilde{b} = (t^T)^{-1} b, \tag{20}$$

where the bets $\tilde{F}$ are uncorrelated, the authors show the diversification distribution to be given by

$$p = \frac{((t^T)^{-1}) \circ (t\Sigma b)}{b^T \Sigma b}. \tag{21}$$

Lohre, Opfer, and Ország (2014) denote the portfolio resulting from maximising (19) as diversified risk parity portfolio (DRP). For (19) to be maximal the diversification distribution has to be uniform. Hence, the analytical optimal portfolio can be found by choosing $\tilde{b}$ in (18) such that $p_i = \frac{1}{k}$ for $i = 1, ..., k$. This holds if

$$\tilde{b}_i = \frac{1}{\sqrt{Var(\tilde{F}_i)}}, \quad i = 1, ..., k. \tag{22}$$

Using (20), the analytical optimal solution can be expressed in terms of the original correlated factors. This solution only holds if no investments constraints are applied. However, many asset managers are restricted to long-only investments, in which case the following optimization has to be solved numerically

$$\omega_{DRP} = \arg\max_{b \in C} N_{Ent}(b), \tag{23}$$

where C is a set of portfolio constraints, such as long-only or full investment constraints.

3.2.2 Principal component versus minimum-torsion bets

While theoretically appealing, the PCA-approach to construct uncorrelated be [Meucci (2009)] is subject to some criticism. Meucci, Santangelo, and Deguest (2015) stress that the approach is purely statistical and might have little relation to the investment process. The ensuing principal portfolios can be difficult to interpret from an economic point of view and are not uniquely determined. Assuming e_k is an eigenvector, then $-e_k$ is also an eigenvector. As a result, there are 2^k possible choices for constructing a maximally diversified portfolio. From a practical perspective, this is a serious concern since using e_k or $-e_k$ translates to either buying or selling a given principal portfolio. In addition, Meucci, Santangelo, and Deguest (2015) point out that the principal portfolios are not invariant under simple scale transformations. Finally, the likely instability of the PCA is expected to translate into unduly high turnover in a related portfolio strategy.

To overcome the drawbacks of principal portfolios, Meucci, Santangelo, and Deguest (2015) suggest another transformation to de-correlate the original factors F. Among all transformations t leading to uncorrelated factors, they choose the transformation $\tilde{t}$ that minimizes the tracking error with respect to the original factors:

$$\tilde{t}_{MT} = \arg\min_{Cr(tF)=I} NTE\{tF|F\} \tag{24}$$

where $Cr()$ denotes the correlation, I the identity matrix, and NTE the multi-entry normalized tracking error

$$NTE(Z|F) = \sqrt{\frac{1}{k}\sum_{k} Var\left(\frac{Z_k - F_k}{\sigma(F_k)}\right)}. \tag{25}$$

Meucci, Santangelo, and Deguest (2015) call this transformation the minimum-torsion transformation. The returns of the minimum-torsion portfolios are given by

$$\tilde{F}_{MT} = \tilde{t}_{MT} F, \tag{26}$$

where the subscript MT refers to minimum-torsion. Meucci, Santangelo, and Deguest (2015) provide an algorithm for solving (24). Expressing the portfolio return in terms of the uncorrelated minimum-torsion portfolios and using (26) leads to:

$$R_p = F^T b = \left(\tilde{t}_{MT}^{-1}\tilde{F}_{MT}\right)^T b = \tilde{F}_{MT}^T \left(\tilde{t}_{MT}^T\right)^{-1} b = \tilde{F}_{MT}^T \tilde{b}_{MT},$$

where $\tilde{b}_{MT} = \left(\tilde{t}_{MT}^T\right)^{-1} b$ are the minimum-torsion exposures (weights), which allow us to compute the diversification distribution and the effective number of bets via (21):

$$p_{MT}(b) = \frac{((\tilde{t}_{MT}^T)^{-1} b) \circ (\tilde{t}_{MT}\Sigma_F b)}{b^T \Sigma_F b}, \quad N_{ent}(b) = \exp\left[-p_{MT}^T \ln(p_{MT}(b))\right] \tag{27}$$

The minimum-torsion approach addresses all concerns regarding principal portfolios. The minimum-torsion portfolios are invariant under a scaling transformation. Moreover, the resulting factors are more stable and have an economic interpretation, since they are designed to be close to the original factors. For the same reason, the buy or sell decision of a given minimum-torsion portfolio is clearly defined.

4 Hierarchical risk parity for multi-asset multi-factor allocations

In this section we focus on examining hierarchical risk parity strategies in the multi-asset multi-factor domain vis-à-vis the alternative risk-based allocation strategies. When it comes to performance comparisons, we make use of various performance measures, including Sharpe ratios and measures that focus on the downside risk potentially plaguing such factor strategies. In evaluating these strategies, we are mindful that their performance is just one realization over the historical path. Therefore, we resort to generate many alternative paths based on block-bootstrapping the historical asset and factor history. We will specifically use the stationary block-bootstrap of Politis and Romano (1994) with an expected block-length of 15 days. This choice helps to preserve stylized facts of asset and factor returns, such as serial correlation and heteroskedasticity.

4.1 Strategy universe

The traditional risk-based allocation strategies are first directly applied to the single assets and factors, and, second, to the eight aggregated factors resulting from the imposed risk model. These eight factors can be viewed as "economic" clusters, providing a benchmark for the "statistical" hierarchical clustering. As for HRP, the allocation strategies used either within or across clusters are risk parity, either based on inverse volatility (IVP) or equal risk contributions (ERC). For hierarchical clustering, we use Ward's method and the dissimilarity matrices, either based on the correlation matrix or the LTDCs. For comparison purposes, two versions of Lopez De Prado's HRP-strategy are considered, both based on recursive bisection: First, we replicate the original strategy, using single linkage and inverse variance allocation as described in Algorithm 1. Second, we consider a variation thereof described in Algorithm 2, using Ward's method and IVP, enabling comparability with the original HRP-strategies of Lopez de Prado. An overview of the considered strategies can be found in Table 1.

Abb.	Description	Formula	Dissimilarity	Linkage	Allocation method	Within and across cluster allocation
Panel A: Strategies based on single assets and factors						
Equal	Equal weighted	(10)				
MVP	Minimum-variance	(11)				
IVP	Inverse volatility	(12)				
ERC	Equal risk contribution	(15)				
Panel B: Strategies based on eight economic factors						
Equal_F	Equal weighted	(10)				
MVP_F	Minimum-variance	(11)				
IVP_F	Inverse volatility	(12)				
ERC_F	Equal risk contribution	(15)				
DRP	Diversified risk parity	(22)				
Panel C: Strategies based on clusters						
DeP_Var	Hierarchical risk parity	Algo. 1	Correlation	Single	Bisection	Inverse variance
DeP_IVP	Hierarchical risk parity	Algo. 1	Correlation	Ward	Bisection	Inverse volatility
H_IVP_C	Hierarchical risk parity	Algo. 2	Correlation	Ward	Cluster	Inverse volatility
H_ERC_C	Hierarchical risk parity	Algo. 2	Correlation	Ward	Cluster	Equal risk contribution
H_IVP_L	Hierarchical risk parity	Algo. 2	LTDC	Ward	Cluster	Inverse volatility
H_ERC_L	Hierarchical risk parity	Algo. 2	LTDC	Ward	Cluster	Equal risk contribution

Table 1: Overview of allocation strategies.

Portfolio rebalancing is conducted on a monthly basis. The strategies are assumed to be implemented using futures and swaps with associated transaction costs of 10 basis points (futures) and 35 basis points (swaps), respectively. Furthermore, eight basis points per month are considered for holding a given swap.

A two-year rolling window of daily returns is used for the estimation of the covariance matrix, and the resulting correlation-based dendrograms are updated every month. As suggested by De Luca and Zuccolotto (2011), we are using an eight-year rolling window to obtain an accurate estimate for the LTDCs. In addition, a univariate Student-t AR(1)-GARCH(1,1) model is applied to each time series to remove autocorrelation and heteroskedasticity. The resulting standardized residuals serve to obtain the pseudo observations, required for the LTDC-estimators.

4.2 A statistical horse race of risk-based allocation strategies

We performed backtests of the investment strategies in the six-year period from January 2012 to December 2017. This rather short sample period resonates with the eight-year estimation window required to determine the LTDCs. To foster intuition regarding the strategies' mechanics, we investigate weight and risk allocations for a subset of the discussed strategies, and Figure 7 shows their weight allocations (left) and volatility contributions by minimum-torsion portfolios (right). The MVP has a fairly high portfolio concentration, allocating up to 20% in individual assets and factors. In the case of the ERC-strategy all assets contribute equally to portfolio variance by design, nonetheless the risk decomposition by minimum-torsion factors looks fairly balanced as well. Finally, the DRP that is designed to maximize diversification w.r.t. these minimum-torsion factors naturally shows a perfectly balanced risk allocation.

Turning to the hierarchical risk parity strategies in Figure 8 we find the original strategy of Lopez de Prado (2016), labeled $DeP\ IVP$, and the other correlation based HRP strategy ($H\ IVP\ C$) to have a higher turnover than the traditional risk parity strategies. Notably, this turnover pattern is less pronounced for the HRP strategy based on lower tail dependencies, suggesting that the corresponding structure is more stable than the one based on correlations. All the three HRP strategies exhibit certain concentrations in terms of the minimum-torsion portfolios, rendering them slightly less diversified than the traditional alternatives.

Table 2 shows performance and risk statistics as well as the average strategy turnover. First, we note that the $1/N$ strategy across single assets and factors has the highest return across all strategies (see Panel A). At the same time, $1/N$ suffers from the highest volatility as well as the highest maximum drawdown, rendering its risk-adjusted performance sub par. Notably, the underlying lack of diversification is not mitigated when considering economic factors as opposed to single assets and factors (Panel B); both variants hover around 3.5 bets averaged over time. Interestingly, minimum-variance optimization enables to already double this number to 7.2 (or 6.7 for the MVP-variant based on economic factors). Unsurprisingly, these two portfolios exhibit the smallest portfolio volatilities in the sample period (0.84% and 0.90%, respectively). Of course, maximum drawdown figures and risk-adjusted returns are also improved relative to equal weighting.

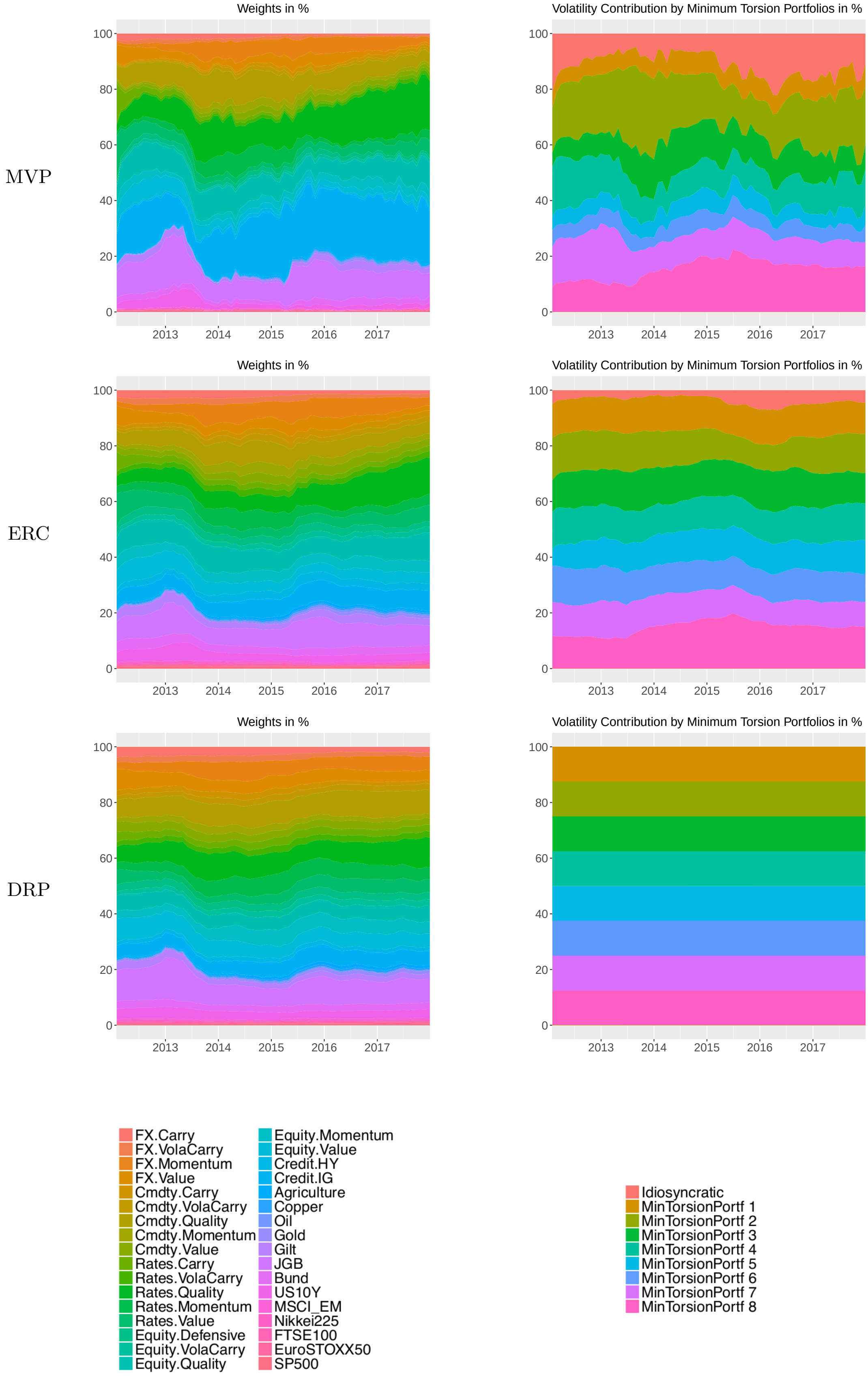

Figure 7: Weight and risk decomposition of selected risk-based allocation strategies

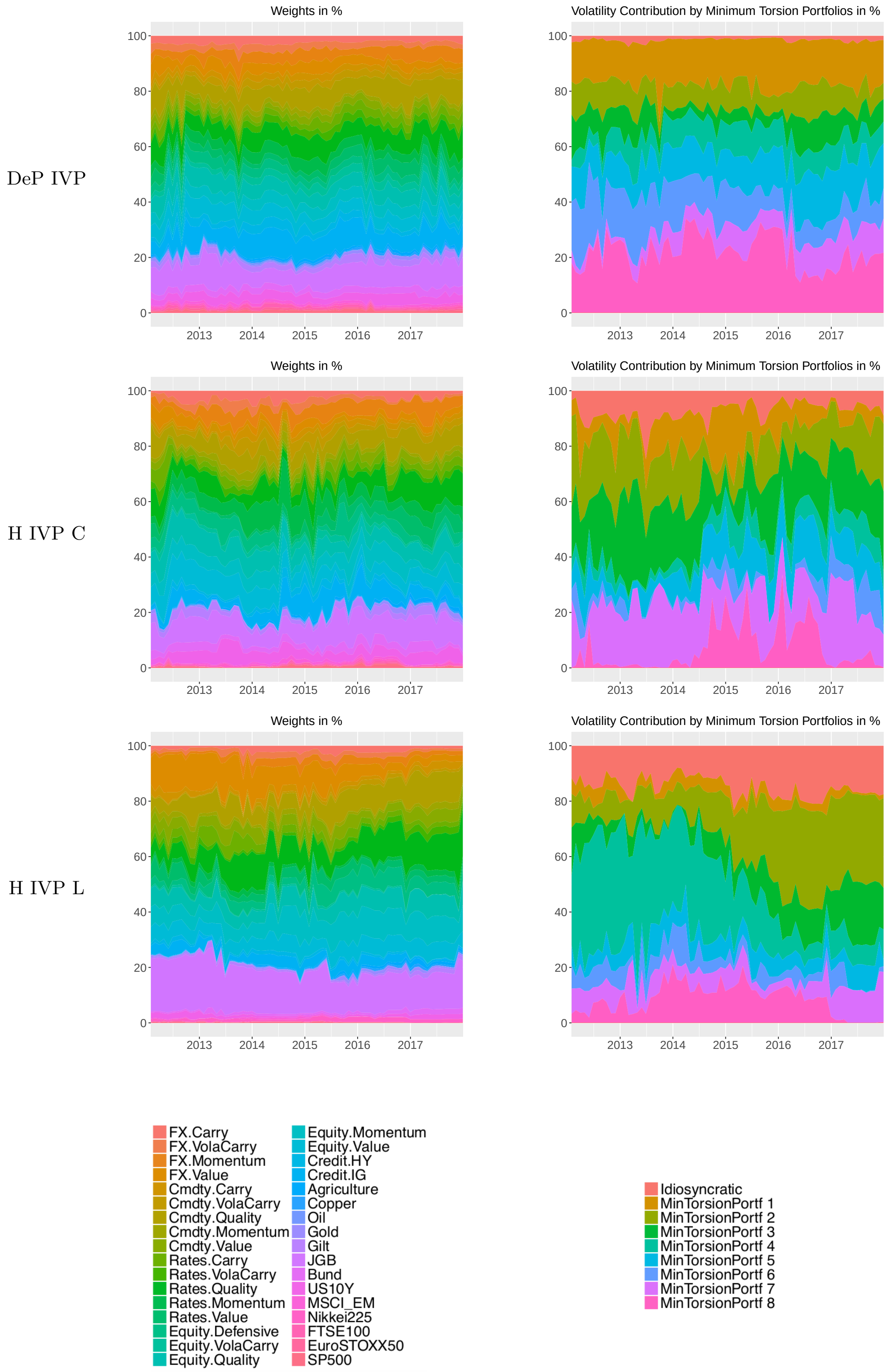

Figure 8: Weight and risk decomposition of hierarchical risk parity strategies

	Risk-Based Allocations (Assets)					
	Equal	MVP	IVP	ERC		
Gross return p.a. in %	3.45	2.37	2.79	2.67		
Net return p.a. in %	2.96	1.61	2.2	1.99		
Volatility p.a. in %	2.84	0.84	1.19	1.00		
Sharpe Ratio	0.86	1.29	1.41	1.46		
Max Drawdown in %	-5.30	-1.62	-1.89	-1.81		
Calmar Ratio	0.53	0.94	1.12	1.05		
Sortino Ratio	1.46	2.83	2.60	2.86		
# of Bets (Avg.)	3.75	7.19	6.06	7.90		
Turnover in % (Avg.)	3.14	11.08	3.34	5.58		
	Risk-Based Allocations (Economic Factors)					
	Equal_F	MVP_F	IVP_F	ERC_F	DRP	
Gross return p.a. in %	2.13	2.32	2.41	2.51	2.57	
Net return p.a. in %	1.59	1.53	1.73	1.83	1.90	
Volatility p.a. in %	1.94	0.95	1.04	1.01	1.00	
Sharpe Ratio	0.56	1.07	1.16	1.30	1.38	
Max Drawdown in %	-4.00	-1.44	-1.51	-1.55	-1.56	
Calmar Ratio	0.34	1.00	1.09	1.12	1.16	
Sortino Ratio	1.16	2.35	2.38	2.62	2.74	
# of Bets (Avg.)	3.54	6.65	7.42	7.95	8.00	
Turnover in % (Avg.)	3.50	5.87	3.65	4.33	4.42	
	Correlation-Based HRP-Allocations				LTDC-Based HRP-Allocations	
	DeP_Var	DeP_IVP	H_IVP_C	H_ERC_C	H_IVP_L	H_ERC_L
Gross return p.a. in %	2.58	2.56	2.29	2.28	2.75	2.71
Net return p.a. in %	1.61	1.55	1.08	1.11	1.71	1.73
Volatility p.a. in %	1.17	1.10	1.14	1.10	1.10	1.08
Sharpe Ratio	0.94	0.94	0.50	0.55	1.08	1.13
Max Drawdown in %	-1.84	-1.65	-1.89	-1.83	-1.20	-1.29
Calmar Ratio	0.82	0.89	0.47	0.52	1.68	1.48
Sortino Ratio	1.91	1.98	1.31	1.40	2.68	2.67
# of Bets (Avg.)	5.94	7.12	6.08	6.71	5.67	6.17
Turnover in % (Avg.)	17.68	18.17	23.21	21.93	15.84	14.14

Table 2: Performance and risk statistics for the multi-asset multi-factor case. All performance figures refer to net returns, except for the gross return. The sample period is from January 2012 to December 2017.

Next, we examine the middle-ground solution in between $1/N$ and minimum-variance: risk parity. In that regard, inverse volatility (ignoring correlations) for single assets and factors shows the second highest return, yet comes at the cost of some diversification. With 6.06 bets over time, it is short around two bets relative to the ERC variant (accounting for correlations). Still, risk-adjusted performance is fairly comparable across the two strategies, and their characteristics are more aligned when switching to economic factors as building blocks. Actually, the corresponding strategy, ERC_F, has fairly similar performance characteristics to the diversified risk parity (DRP) strategy that is designed to have maximum number of eight bets throughout time.

Having investigated the risk-based strategies for economic factors, we are eager to learn how the approaches based on statistical clusters fare. We start off by examining the original strategy of Lopez de Prado (2016). Its gross return is slightly higher than the one of the risk parity variants, yet its turnover is more than three times higher (∼18% versus ∼5%). As a result, net returns are considerably smaller as well as net Sharpe ratios (0.94 versus 1.26 for ERC_F). Interestingly, the two further correlation-based HRP allocations (H_IVP_C and H_ERC_C) have even higher turnover (23.21% and 21.93%, respectively), bringing net Sharpe ratios down to 0.50 and 0.55.

Finally, we investigate the effect of replacing the correlation-based dissimilarity matrix by one that is driven by LTDCs. We would hope for improved tail risk statistics of related strategies, and indeed, we observe H_IVP_L and H_ERC_L to experience the two smallest maximum drawdowns over the sample period, -1.20% and -1.29%, respectively. These compare to the third smallest value for MVP_F (-1.44%) and the fourth smallest value for IVP_F (-1.51%). The two LTDC-based HRP strategies' turnover is less elevated, rendering these two strategies more competitive relative to the alternative risk parity strategies in terms of (risk-adjusted) returns.

Of course, all of the above evidence might be considered anecdotal as they merely refer to the realization over the historical path (which is also rather short). We wish to foster intuition with regard to the distributional properties of the obtained performance statistics. Using the 2003–2017 data history, we generated 2,000 "alternative" six-year investment periods based on the stationary block-bootstrap of Politis and Romano (1994) with an average block length of 15 days.[2] Figure 9 shows the boxplots of the associated performance and risk statistics. Note that we have dropped Equal and Equal_F for not distorting the charts.

By and large, we find the block-bootstrapped results to be consistent with the historical evidence. Still, it seems that the correlation-based HRP strategies have higher median net returns, even though their turnover is higher than what was observed historically. The corresponding dendrograms' instability is crucially driving the high turnover. In total, the general return ranking is unchanged and we note that strategies based on single assets and factors have wider return distributions than those based on economic factors. These findings also carry over to net Sharpe ratios. A further divergence from the historical evidence is that the median maximum drawdowns of the HRP strategies based on LTDCs are much lower than the maximum drawdown results of strategies based on economic factors. Hence, the observed backtest advantage might be deemed more apparent than real. Against this backdrop, the ranking of downside risk-adjusted returns (Calmar ratio) are more aligned along strategies either operating across economic factors or statistical clusters.

[2]Given the computational effort of fitting the AR-GARCH models, the LTDC-based dendrogram is only updated once a year for the purpose of the block-bootstrap analysis. This yearly updating of the dendrogram reduces the strategy turnover considerably, yet it likewise reduces the strategy's ability to timely adjust to changes in the risk environment. These two effects tend to cancel on average, and unreported results document the corresponding historical backtest to be consistent in terms of net returns with the one based on monthly updating.

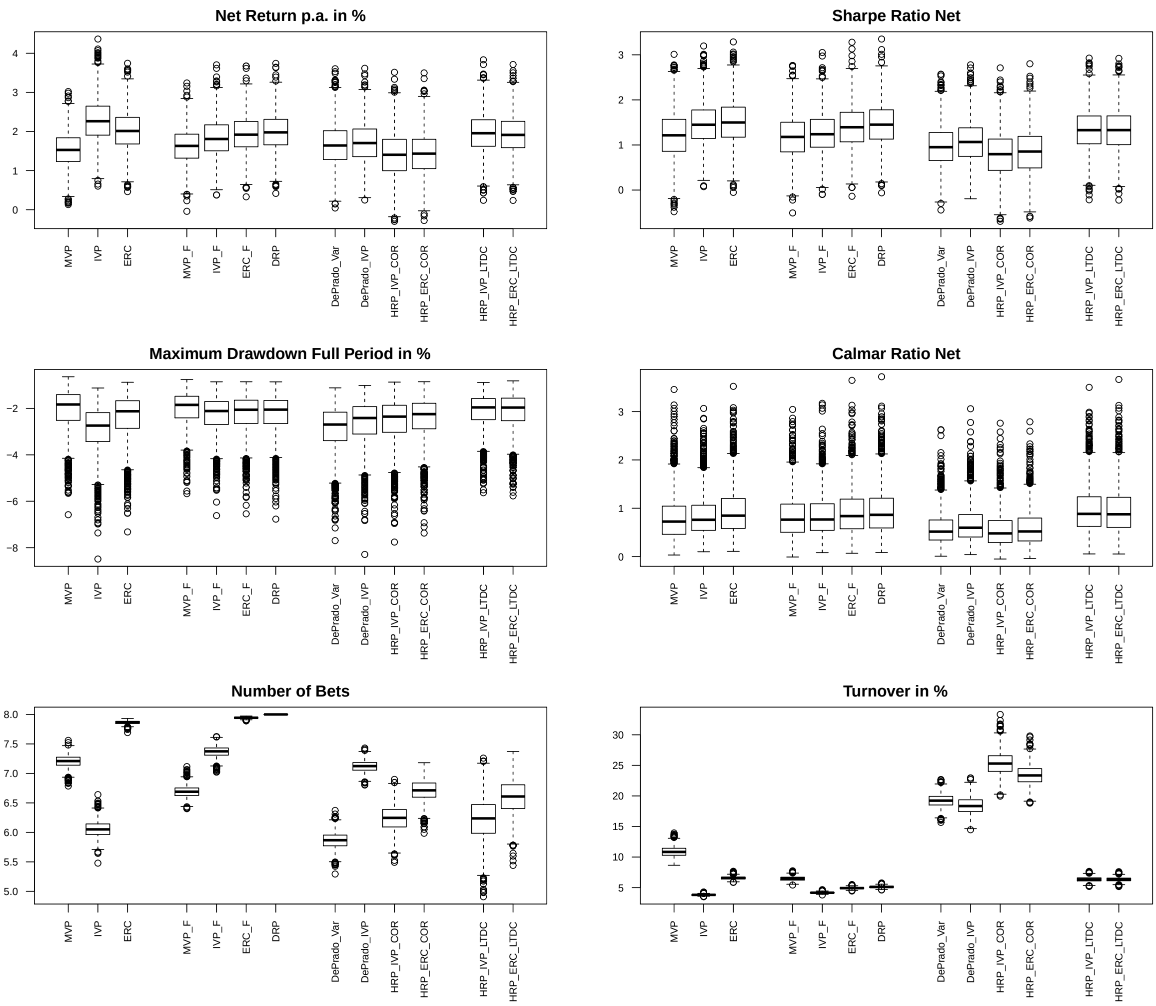

Figure 9: Boxplots of performance and risk statistics for a six-year investment horizon, based on 2,000 block-bootstrap simulations.

5 Conclusion

Building upon hierarchical risk parity of Lopez de Prado (2016), we proposed and investigated innovative variations of this strategy. The main motivation to base an allocation strategy on hierarchical clustering is that the correlation matrix is too complex to be fully analyzed and lacks the notion of hierarchy. Hierarchical clustering allows us to reduce the associated complete graph of information into a minimum spanning tree, thus focusing on the correlations that really matter. Hierarchical risk parity is an intuitive investment approach, allowing for a high degree of flexibility. We can choose from various dissimilarity measures, trying to capture different properties and relations among the underlying return time series. Moreover, we can modify the resulting tree structure by considering various linkage methods. The only restriction is to impose that weights of within- and across-cluster allocation strategies each sum up to 1.

As an alternative to the common choice of the Pearson correlation coefficient, we build a dissimilarity measure based on the lower tail dependence to explicitly incorporate the notion of tail risk management. Such innovation is expected to help navigate the multi-asset multi-factor universe that comes with significant tail risk for some of its constituents. Having obtained the hierarchical structure of assets and factors in terms of the dendrogram, a simple and efficient capital allocation is applied within and across clusters. Lopez de Prado (2016) suggested to use an algorithm based on recursive bisection, which only considers the order of the assets at the "bottom" of the dendrogram and ignores the nested structure of the clusters. Conversely, we defined an allocation algorithm that fully uses the hierarchical structure by allocating capital in a top-to-bottom fashion according to the clusters created at each level of the dendrogram.

A multi-asset multi-factor case study suggests that hierarchical risk parity strategies have the potential to compete with traditional risk parity strategies. Moreover, they can add desirable diversification properties. Especially, the hierarchical risk parity strategies based on the lower tail dependence coefficient seem to benefit tail risk management. Notably, traditional risk parity strategies based on economic factors can be expected to provide downside risk protection that is on par with these HRP strategies. Thus, diversifying across economic factor clusters is meaningful as well. Yet, higher-dimensional investment universes could make the "economic" clustering less intuitive and thus more difficult to come up with reasonable clusters. In contrast, ignoring computational intensity, hierarchical clustering methods can be directly applied to high-dimensional investment universes.

References

Bernardi, S., M. Leippold, and H. Lohre (2018): "Maximum diversification strategies along commodity risk factors," *European Financial Management*, 24(1), 53–78.

Boudt, K., P. Carl, and B. Peterson (2013): "Asset allocation with conditional value-at-risk budgets," *Journal of Risk*, 15(3), 39–68.

De Luca, G., and P. Zuccolotto (2011): "A tail dependence-based dissimilarity measure for financial time series clustering," *Advances in Data Analysis and Classification*, 5(4), 323–340.

Durante, F., R. Pappadà, and N. Torelli (2015): "Clustering of time series via non-parametric tail dependence estimation," *Statistical Papers*, 56(3), 701–721.

Jurczenko, E., and J. Teiletche (2019): "Expected Shortfall Asset Allocation: A Multi-Dimensional Risk Budgeting Framework," *Journal of Alternative Investments*, 22(Fall), 1–16.

Lohre, H., H. Opfer, and G. Ország (2014): "Diversifying risk parity," *Journal of Risk*, 16(5), 53–79.

Lopez de Prado, M. (2016): "Building diversified portfolios that outperform out of sample," *Journal of Portfolio Management*, 42(4), 59–69.

Maillard, S., T. Roncalli, and J. Teiletche (2010): "The properties of equally weighted risk contribution portfolios," *Journal of Portfolio Management*, 36(4), 60–70.

Mantegna, R. (1999): "Hierarchical structure in financial markets," *The European Physical Journal B - Condensed Matter and Complex Systems*, 11(1), 193–197.

Markowitz, H. (1952): "Portfolio Selection," *Journal of Finance*, 7(1), 77–91.

Meucci, A. (2009): "Managing diversification," *Risk*, pp. 35–40.

Meucci, A., A. Santangelo, and R. Deguest (2015): "Risk budgeting and diversification based on optimized uncorrelated factors," *Risk*, 11(29), 70–75.

Papenbrock, J. (2011): "Asset Clusters and Asset Networks in Financial Risk Management and Portfolio Optimization," Ph.D. thesis.

Politis, D. N., and J. P. Romano (1994): "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89(428), 1303–1313.

Prim, R. C. (1957): "Shortest Connection Networks And Some Generalizations," *Bell System Technical Journal*, 36(6), 1389–1401.

Raffinot, T. (2017): "Hierarchical Clustering-Based Asset Allocation," *Journal of Portfolio Management*, 44(2), 89–99.

Rousseeuw, P. J. (1987): "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, 20(C), 53–65.

Schmid, F., and R. Schmidt (2007): "Multivariate conditional versions of Spearman's rho and related measures of tail dependence," *Journal of Multivariate Analysis*, 98(6), 1123–1140.

Tibshirani, R., G. Walther, and T. Hastie (2001): "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 411–423.

Tumminello, M., F. Lillo, and R. N. Mantegna (2010): "Correlation, hierarchies, and networks in financial markets," *Journal of Economic Behavior and Organization*, 75(1), 40–58.

A Definition of style factors

A.1 Foreign exchange (FX) style factors

Data	Style	Description
GS	Carry	The FX Carry strategy has historically benefited from the tendency of FX forwards of high yielding currencies to overestimate the actual depreciation of future FX spot. On a monthly basis, the strategy evaluates the implied carry rate (FX Forward vs FX Spot) of a number of currencies (G10 and EM) against the USD and ranks them based on that measure. The strategy goes long single currency indices (which roll FX forwards) for the currencies with the highest carry and short single currency indices for the currencies with the lowest carry.
GS	Value	The FX Valuation strategy relies on exchange rates reverting back to their fair value in the medium to long-term horizons. On a monthly basis, the strategy evaluates the valuation measure (based on GS DEER, Dynamic Equilibrium Exchange Rate model) of a number of currencies (G10 and EM) against the USD and ranks them based on that measure. The strategy goes long single currency indices (which roll FX forwards) for the currencies with the highest ranking (i.e. most undervalued) and short single currency indices for the currencies with the lowest ranking (i.e. most overvalued).
GS	Momentum	This factor capitalizes on the persistence of trends in forward exchange rate movements which are driven both by carry as well as spot movements. On a daily basis, the strategy evaluates the recent performance of 27 currencies against the USD. It then takes either a long or short position on each of those currencies against the USD, depending on whether their actual performance has been positive or negative.
GS	Vola	On average, implied volatility tends to trade at a premium to subsequent realized volatility as investors demand a risk premium for selling optionality and being short volatility. This preference is driven by risk aversion, i.e. market participants attach a higher value to being long protection. FX Volatility Carry sells short-dated options (puts and calls) on several currencies and delta-hedges on a daily basis.

A.2 Commodity style factors

Data	Style	Description
GS	Carry	Captures tendency for commodities with tighter timespreads to outperform due to low inventories driving both backwardated futures curves and price appreciation, and buying demand from consumer hedgers for protection against price spikes in undersupplied commodities. The strategy goes long the top third and short the bottom third of the 24 commodities from the S&P GSCI universe, ranked by annualized strength of front month time spreads. The strategy is rebalanced daily based on the signals over the last 10 days. The strategy is net of cost.
GS	Value	The strategy uses the weekly Commodity Futures Trading Commission (CFTC) positioning data to determine to go long and short in commodities and will take long positions in the commodities that the cumulative positions are most short and short position in the commodities that speculative positions are most long.
GS	Momentum	Momentum in commodity returns reflect initial underreaction or subsequent overreaction to changes in demand as increasing or decreasing supply take many years to implement and subsequently overshoot required changes to match demand. The strategy goes long the top third and short the bottom third of the 24 commodities from the S&P GSCI universe, ranked by rolling 1-year excess returns of each commodity. The strategy is rebalanced daily based on the signals over the last ten days. The strategy is net of cost.
GS	Quality	This factor captures the tendency for deferred futures contracts to outperform nearer dated contracts due to producers hedging further out than consumers, and passive investors investing near the front of the curve. The strategy goes long selected points on the curve of each commodity, equally weighted amongst commodities. The strategy goes short an equally weighted basket of the nearest commodity contracts, beta-adjusted at the basket level.
GS	Vola	Options on commodities are used by hedgers and investors to protect against falling or rising prices, or to speculate with limited downside risk. Such participants are generally buyers of implied volatility. Therefore the strategy aims to capture the risk premium of convexity transfer.

A.3 Rates style factors

Data	Style	Description
GS	Momentum	This factor capitalizes on the persistence of trends in short and long-term interest rate movements. On a daily basis, the strategy evaluates the recent performance of a number of futures contracts for US, Germany, Japan and UK. It then takes either a long or short position on each of the futures, depending on whether their actual performance has been positive or negative.
GS	Quality	This factor capitalizes on the observation that risk-adjusted returns at the short-end of the curve tend to be higher than at the long-end. A leveraged long position on the former vs the latter tends to capture positive excess returns as compensation for the risk premium that stems from investors having leverage constraints and favouring long-term rates. The Interest Rates Curve strategy enters a long position on 5y US Bond futures and a short position on 30y Bond futures, as well as a long position on 5y German Bond futures and a short position on 10y German Bond futures, rolling every quarter. The exposure to each future is adjusted to approximate a duration-neutral position.
GS	Vola	On average, Treasury Futures implied volatility tends to trade at a premium to subsequent realized volatility as investors demand a risk premium for selling optionality and being short volatility. This preference is driven by risk aversion, i.e. market participants attach a higher value to being long protection. Interest Rates Volatility Carry sells short-dated options (puts and calls) monthly on 10y Treasury Futures and delta-hedges on a daily basis.

A.4 Equity style factors

Data	Style	Description
IQS	Value	The Value factor refers to the finding that value stocks characterized by valuation ratios offer higher long-run average returns than growth stocks characterized by high valuation ratios. To optimally combine factors in a multi-factor asset allocation exercise we construct the factor to have zero exposure to other factors and thus does not alter the overall portfolio exposure to other factors when added to the portfolio.
IQS	Momentum	The Momentum factor captures a medium-term continuation effect in returns by buying recent winners and selling recent losers. The factor combines price as well as earnings momentum information. To optimally combine factors in a multi-factor asset allocation exercise we construct the factor to have zero exposure to other factors and thus does not alter the overall portfolio exposure to other factors when added to the portfolio.
IQS	Quality	The Quality factor combines different measures of determining financial health and operating profitability. To optimally combine factors in a multi-factor asset allocation exercise we construct the factor to have zero exposure to other factors and thus does not alter the overall portfolio exposure to other factors when added to the portfolio.
IQS	Defensive	The Defensive factor refers to the finding that low volatility stocks tend to outperform high volatile stocks on a risk-adjusted basis. To capture this behaviour the factor is constructed to go long a minimum-variance portfolio and short the beta-portion of the market
GS	Vola	On average, equity implied volatility tends to trade at a premium to subsequent realized volatility as investors demand a risk premium for selling optionality and being short volatility. This preference is driven by risk aversion, i.e. market participants attach a higher value to being long protection. Global Equity Vol Carry sells short-dated options (puts and calls) daily on global indexes and delta-hedges. It buys back lower puts to mitigate drawdowns without diluting the underlying alpha drivers.

B CSR-Estimator

We state the estimator as presented in Schmid and Schmidt (2007). Consider a random sample $(X_j)_{j=1,...,n}$ from a d-dimensional vector X with joint distribution function F and copula C. Based on the multivariate Conditional version of Spearman's Rho (CSR), given by

$$\rho(p) = \frac{\int_{[0,p]^d} C(u)du - \left(\frac{p^2}{2}\right)^d}{\frac{p^{d+1}}{d+1} - \left(\frac{p^2}{2}\right)^d}, \quad 0 < p \leq 1, \tag{28}$$

they define a multivariate LTDC ρ_L, a generalization of the bivariate LTDC λ_L, by

$$\rho_L = \lim_{p \searrow 0} \rho(p) = \lim_{p \searrow 0} \frac{d+1}{p^{d+1}} \int_{[0,p]^d} C(u)du. \tag{29}$$

Assuming that neither the univariate distribution functions F_{X_j} nor the according copula C are known, empirical versions have to be considered. An empirical estimator for the marginal distributions is given by:

$$\hat{F}_{i,n}(x) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{\{X_{ij} \leq x\}}, \quad \text{for } i = 1, ..., d \text{ and } x \in \mathbb{R}. \tag{30}$$

Set $\hat{U}_{ij,n} = \hat{F}_{i,n}(X_{ij})$ for $i = 1, ..., d$, $j = 1, ..., n$ and $\hat{U}_{j,n} = (\hat{U}_{1j,n}, ..., \hat{U}_{dj,n})$. Note that

$$\hat{U}_{ij,n} = \frac{1}{n} \left(\text{rank of } X_{ij} \text{ in } X_{i1}, ..., X_{in}\right). \tag{31}$$

The copula C is estimated by the empirical copula:

$$\hat{C}_n(u) = \frac{1}{n} \sum_{j=1}^{n} \prod_{i=1}^{d} \mathbb{1}_{\{\hat{U}_{ij,n} \leq u_i\}} \quad \text{for } u = (u_1, ..., u_d)^T \in [0,1]^d. \tag{32}$$

Leading to the following estimator for (28):

$$\hat{\rho}_n(p) = \left\{ \frac{1}{n} \sum_{j=1}^{n} \prod_{i=1}^{d} \left(p - \hat{U}_{ij,n}\right)^+ - \left(\frac{p^2}{2}\right)^d \right\} \Big/ \left\{ \frac{p^{d+1}}{d+1} - \left(\frac{p^2}{2}\right)^d \right\} \tag{33}$$

Finally, the estimator for (29) is given by:

$$\hat{\rho}_L = \hat{\rho}_n\left(\frac{k}{n}\right), \tag{34}$$

where $k \in \{1, ..., n\}$ has to be chosen by the statistician.