The Journal of Portfolio Management

Hierarchical Clustering-Based Asset Allocation

Thomas Raffinot

JPM 2017, 44 (2) 89-99
doi: https://doi.org/10.3905/jpm.2018.44.2.089
http://jpm.iijournals.com/content/44/2/89

Hierarchical Clustering-Based Asset Allocation

Thomas Raffinot

Thomas Raffinot
is the Head of Quantitative Macro Strategy at Silex-IP in Paris, France.
traffinot@gmail.com

Nobel Prize winner Harry Markowitz described diversification, with its ability to enhance portfolio returns while reducing risk, as the "only free lunch" in investing (Markowitz [1952]). Yet, diversifying a portfolio in real life is easier said than done.

Investors are aware of the benefits of diversification but form portfolios without giving proper consideration to the correlations (Goetzmann and Kumar [2008]). Moreover, modern and complex portfolio optimization methods are optimal in sample but often provide rather poor out-of-sample forecast performance. For instance, DeMiguel et al. [2009] demonstrate that the equal-weighted allocation, which gives the same importance to each asset, beats the entire set of commonly used portfolio optimization techniques. In fact, optimized portfolios depend on expected returns and risks, but even small estimation errors can result in large deviations from optimal allocations in an optimizer's result (Michaud [1989]).

To overcome this issue, academics and practitioners have developed risk-based portfolio optimization techniques (minimum variance, equal-risk contribution, risk budgeting, etc.) that do not rely on return forecasts (Roncalli [2013]). However, these still require the inversion of a positive-definite covariance matrix, which leads to errors of such magnitude that they entirely offset the benefits of diversification (López de Prado [2016b]).

Exploring a new way of capital allocation, López de Prado [2016a] introduces a portfolio diversification technique called hierarchical risk parity (HRP). One of the main advantages of HRP is in computing a portfolio on an ill-degenerated or even a singular covariance matrix. Lau et al. [2017] apply HRP to different cross-asset universes consisting of many tradable risk premia indexes and confirm that HRP delivers superior risk-adjusted returns. Alipour et al. [2016] propose a quantum-inspired version of HRP, which outperforms HRP and thus other conventional methods.

The starting point of HRP is that a correlation matrix is too complex to be properly analyzed and understood. If you have N assets of interest, there are $\frac{1}{2}N(N-1)$ pairwise correlations among them and that number grows quickly. For example, there are as many as 4,950 correlation coefficients between stocks of the FTSE 100 and 124,750 between stocks of the S&P 500. More importantly, correlation matrices lack the notion of hierarchy. Actually, Nobel Prize laureate Herbert Simon has argued that complex systems can be arranged in a natural hierarchy comprising nested substructures (Simon [1962]). But, a correlation matrix makes no differentiation between assets. Yet, some assets seem closer substitutes of one another, while others seem complementary to one another. This lack of hierarchical structure allows weights to vary freely in unintended ways (López de Prado [2016a]).

To simplify the analysis of the relationships between this large group of relative prices, López de Prado [2016a] applies a correlation-network method known as the "minimum spanning tree (MST)."[1] Its main principle is easy to understand: the heart of correlation analysis is choosing which correlations really matter; in other words, choosing which links in the network are important and removing the rest, keeping $N - 1$ links.

Graph theory is linked to unsupervised machine learning. For instance, the MST is strictly related to a hierarchical clustering algorithm, named the *single linkage* (Tumminello et al. [2010]). Hierarchical clustering refers to the formation of a recursive clustering, suggested by the data, not defined a priori. The objective is to build a binary tree of the data that successively merges similar groups of points. Hierarchical clustering is thus another way to filter correlations.

Finally, there are many generalizations of the MST. The *planar maximally filtered graph* (Tumminello et al. [2005]) is a recent and prominent one. It is associated with a hierarchical clustering method, the directed bubble hierarchical tree (DBHT) (Musmeci et al. [2015]).

Building upon López de Prado [2016a] and Simon [1962], this article exploits the notion of hierarchy. Different hierarchical clustering methods are presented and tested, namely, simple linkage, complete linkage, average linkage, Ward's method, and DBHT. Once the assets are hierarchically clustered, a simple and efficient capital allocation within and across clusters of assets at multiple hierarchical levels is computed.

The out-of-sample performances of hierarchical clustering-based portfolios and risk-based portfolios are evaluated across three empirical datasets, which differ in terms of number of assets and composition of the universe (S&P sectors, multi-assets, and individual stocks). To avoid data snooping, which occurs when a given set of data is used more than once for purposes of inference or model selection, the comparison of profit measures is assessed using the bootstrap-based model confidence set procedure proposed by Hansen et al. [2011]. It prevents strategies that perform by luck to be considered as effective.

The findings of this article can be summarized as follows: hierarchical clustering-based portfolios are robust, truly diversified, and achieve statistically better risk-adjusted performances than commonly used portfolio optimization techniques. Among clustering methods, there is no clear winner. DBHT-based portfolios produce slightly superior risk-adjusted returns, but average-linkage-based portfolios are clearly more robust.

HIERARCHICAL CLUSTERING AND ASSET ALLOCATION

Notion of Hierarchy

Nobel Prize winner Herbert Simon has argued that complex systems, such as financial markets, have a structure and are usually organized in a hierarchical manner, with separate and separable substructures (Simon [1962]). The hierarchical structure of interactions among elements strongly affects the dynamics of complex systems. The need of a quantitative description of hierarchies to model complex systems is thus straightforward (Anderson [1972]).

López de Prado [2016a] points out that correlation matrices lack the notion of hierarchy, which allows weights to vary freely in unintended ways. He provides a concrete example to highlight the interest of the notion of hierarchy for asset allocation:

> *Stocks could be grouped in terms of liquidity, size, industry, and region, where stocks within a given group compete for allocations. In deciding the allocation to a large publicly-traded U.S. financial stock like J.P. Morgan, we will consider adding or reducing the allocation to another large publicly-traded U.S. bank like Goldman Sachs, rather than a small community bank in Switzerland, or a real estate holding in the Caribbean.*
>
> —López de Prado [2016a]

To sum up, a correlation matrix makes no differentiation between assets. Yet, some assets seem to be closer substitutes of one another, while others seem to be complementary to one another.

Hierarchical Clustering

The purpose of cluster analysis is to place entities into groups, or clusters, suggested by the data, not defined a priori, such that entities in a given cluster tend to be similar to each other and entities in different clusters tend to be dissimilar.

Hierarchical clustering refers to the formation of a recursive clustering. The objective is to build a binary tree of the data that successively merges similar groups of points. The tree-based representation of the observations is called a *dendrogram*. Visualizing this tree provides a useful summary of the data.

Hierarchical clustering requires a suitable distance measure. The following distance is used (Mantegna [1999]):

$$D_{i,j} = \sqrt{2(1-\rho_{i,j})} \tag{1}$$

where $D_{i,j}$ is the correlation-distance index between the *ith* and *jth* asset, and $\rho_{i,j}$ is the respective Pearson's correlation coefficient.

Four agglomerative clustering variants are tested in this study—namely, single linkage (SL), complete linkage (CL), average linkage (AL), and Ward's method (WM).

An agglomerative clustering starts with every observation representing a singleton cluster and then combines the clusters sequentially, reducing the number of clusters at each step until only one cluster is left. At each of the $N-1$ steps, the closest two (least dissimilar) clusters are merged into a single cluster, producing one less cluster at the next higher level. Therefore, a measure of dissimilarity between two clusters must be defined, and different definitions of the distance between clusters can produce radically different dendrograms. The clustering variants are described below:

- *SL*: the distance between two clusters is the minimum of the distance between any two points in the clusters. For clusters C_i, C_j,

$$d_{C_i,C_j} = min_{x,y}\{D(x,y) \mid x \in C_i, y \in C_j\} \tag{2}$$

  This method is relatively simple and can handle nonelliptical shapes. Nevertheless, it is sensitive to outliers and can result in a problem called *chaining*, whereby clusters end up being long and straggly. The SL algorithm is strictly related to the one that provides an MST. However the MST retains some information that the SL dendrogram throws away.

- *CL*: the distance between two clusters is the maximum of the distance between any two points in the clusters. For clusters C_i, C_j,

$$d_{C_i,C_j} = max_{x,y}\{D(x,y) \mid x \in C_i, y \in C_j\} \tag{3}$$

  This method tends to produce compact clusters of similar size but is quite sensitive to outliers.

- *AL*: the distance between two clusters is the average of the distance between any two points in the clusters. For clusters C_i, C_j,

$$d_{C_i,C_j} = mean_{x,y}\{D(x,y) \mid x \in C_i, y \in C_j\} \tag{4}$$

  This is considered to be a fairly robust method.

- *WM* (Ward [1963]): the distance between two clusters is the increase of the squared error that results when two clusters are merged. For clusters C_i, C_j with sizes m_i, m_j, respectively,

$$d_{C_i,C_j} = \frac{m_i m_j}{m_i + m_j} \|c_i - c_j\|^2 \tag{5}$$

  where c_i, c_j are the centroids for the clusters.

This method is biased toward globular clusters but less susceptible to noise and outliers. It is one of the most popular methods.

To determine the number of clusters, we employ the Gap index (Tibshirani et al. [2001]). It compares the logarithm of the empirical within-cluster dissimilarity and the corresponding one for uniformly distributed data, which is a distribution with no obvious clustering.

The last approach differs completely from the agglomerative one: the idea of DBHT is to use the hierarchy hidden in the topology of a planar maximally filtered graph (PMFG) (Tumminello et al. [2005]).

The PMFG network keeps the hierarchical structure of the MST network but contains a greater amount of information by connecting N nodes (assets) with $3(N-2)$ edges. The basic elements of a PMFG are three-cliques (subgraphs made of three nodes all reciprocally connected). For a detailed introduction of MST and PMFG, see Tumminello et al. [2005] and Aste et al. [2010].

Musmeci et al. [2015] explains that the DBHT exploits this topological structure, and in particular, the distinction between separating and nonseparating three-cliques, to identify a clustering partition of all nodes in the PMFG. First of all, the clusters are identified by means of topological considerations on the planar graph, then the hierarchy is constructed both between clusters and within clusters. Therefore, the difference involves both the kind of information exploited and the methodological approach. Note that the "optimal" number of clusters is determined during the process (see Song et al. [2012] for more on this subject).

Asset Allocation Weights

Once the clusters have been determined, the capital should be efficiently allocated both within and across groups. Indeed, a compromise between diversification across all investments and diversification across clusters of investments at multiple hierarchical levels has to be found.

Because asset allocation within and across clusters can be based on the same or different methodologies, there are countless options.

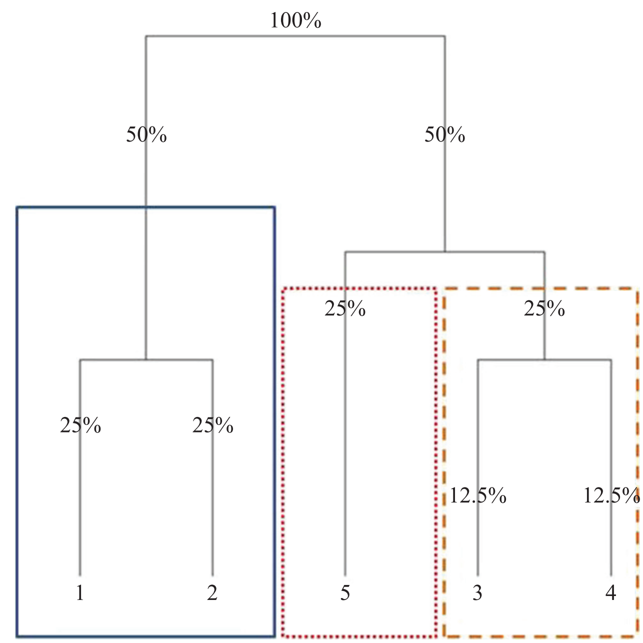
The chosen weighting scheme attempts to stay very simple and focuses not only on the clusterings but also on the entire hierarchies associated with those clusterings. The principle is to find a diversified weighting by distributing capital equally to each cluster hierarchy, so that many correlated assets receive the same total allocation as a single uncorrelated one. Then, within a cluster, an equal-weighted allocation is computed.

For example, Exhibit 1 illustrates a small dendrogram with five assets and three clusters. The first cluster is made up of Assets 1 and 2; Asset 5 constitutes the second cluster, and the third cluster consists of Assets 3 and 4. Based on the hierarchical clustering weighting, weights for cluster 1 is $0.5(\frac{1}{2} = 0.5)$ and weights for clusters 2 and 3 are $0.25(\frac{0.5}{2} = 0.25)$. Because there are two assets in Cluster 1, final weights for Assets 1 and 2 are $\frac{0.5}{2} = 0.25$. Asset 5 would have a weight of $\frac{0.25}{1} = 0.25$. Finally, Assets 3 and 4 would get a weight of $\frac{0.25}{2} = 0.125$.

This weighting scheme should guarantee the diversification and the robustness of the portfolio. For instance, because we consider at least two clusters, the weights are constrained: $\forall i$: $0 \leq w_i \leq 0.5$. Moreover, if clusters are lasting, the weights should be very stable. Finally, neither expected returns nor risk measures are required, thereby making the method more robust.

Exhibit 1

Asset Allocation Weights: ***A Small Example***



The will to exploit the nested clusters or, in other words, the notion of hierarchy explains why clustering methods such as K-means or K-medoids have not been tested. Indeed, these algorithms provide a single set of clusters with no particular organization or structure within them.[2]

RISK-BUDGETING APPROACH

This section briefly describes risk-budgeting portfolios. Refer to Roncalli [2013] for a detailed exposition of this approach. In a risk-budgeting approach, the investor only chooses the risk repartition between assets of the portfolio, without any consideration of returns, thereby partially dealing with the issues of traditional portfolio optimization methods.

Notations and Definitions

Consider a portfolio invested in N assets with portfolio weights vector $w = (w_1, w_2, \ldots, w_N)'$. Returns are assumed to be arithmetic: $r_{t,i} = (p_{t,i} - p_{t-1,i})/p_{t-1,i} = p_{t,i}/p_{t-1,i} - 1$. The portfolio return at time t is thus

$$r_{P,t} = \sum_{i=1}^{N} w_i r_{t,i} \tag{6}$$

Let σ_i^2 be the variance of asset i, σ_{ij} be the covariance between assets i and j, and Σ be the covariance matrix.

The volatility is defined as the risk of the portfolio:

$$R_w = \sigma_w = \sqrt{w'\Sigma w} \tag{7}$$

and μ is the expected return:

$$\mu = E(r_P) = \sum_{i=1}^{N} w_i E(r_i) \tag{8}$$

Risk-Budgeting Portfolios

In a risk-budgeting portfolio, the risk contribution from each components is equal to the budget of risk defined by the portfolio manager.

Because the risk measure is coherent and convex, the Euler decomposition is verified:

$$R_w = \sum_{i=1}^{N} w_i \frac{\partial R_w}{\partial w_i} \tag{9}$$

With the volatility as the risk measure, the risk contribution of the *ith* asset becomes

$$RC_{w_i} = w_i \frac{(\Sigma w)_i}{\sqrt{w'\Sigma w}} \tag{10}$$

A long-only, full invested risk-budgeting portfolio is defined as follows (Roncalli [2013]):

$$\begin{cases} RC_{w_i} = b_i RC_w \\ b_i > 0 \\ \sum_{i=1}^{N} b_i = 1 \\ w_i \geq 0 \\ \sum_{i=1}^{N} w_i = 1 \end{cases} \tag{11}$$

Once a set of risk budgets is defined, the weights of the portfolio are computed so that the risk contributions match the risk budgets.

In this article, four risk-budgeting portfolios are considered:[3]

- The minimum-variance (MV) portfolio is a risk-budgeting portfolio in which the risk budget is equal to the weight of the asset:

$$b_i = w_i \tag{12}$$

- The most diversified portfolio (MDP) (Choueifaty et al. [2013]) is a risk-budgeting portfolio in which the risk budgets are linked to the product of the weight of the asset and its volatility:

$$b_i = \frac{w_i \sigma_i}{\sum_{i=1}^{N} w_i \sigma_i} \tag{13}$$

- The equal-risk-contribution portfolio (ERC) (Maillard et al. [2010]) is a risk-budgeting portfolio in which the risk contribution from each asset is made equal:

$$b_i = \frac{1}{N} \tag{14}$$

- The inverse-variance (IVRB) risk-budgeting portfolio defines risk budgets as follows:

$$b_i = \frac{\sigma_i^{-2}}{\sum_{i=1}^{N} \sigma_i^{-2}} \tag{15}$$

We use the cyclical coordinate descent (CCD) algorithm for solving high-dimensional risk parity problems (Griveau-Billion et al. [2013]) to estimate the risk-based models.

INVESTMENT STRATEGIES COMPARISON

Portfolios are updated on a daily basis via a 252-day rolling-window approach, with no forward-looking biases. This approach differs from the traditional one, in which portfolios are rebalanced on a more realistic monthly basis.[4] Nevertheless, the main objective of this article is not to create a real investment strategy but to compare asset allocation methods. The daily rebalancing framework should help highlight the strengths and weaknesses of the different approaches, especially the robustness.

Datasets

The out-of-sample performances of models are evaluated across three very disparate datasets. The three considered datasets differ in term of assets' composition and number of assets:[5]

- The S&P sectors dataset consists of daily returns on 10 value-weighted industry portfolios formed by using the Global Industry Classification Standard (GICS) developed by Standard & Poor's. The 10 industries considered are energy, material, industrials, consumer-discretionary, consumer-staples, healthcare, financials, information-technology, telecommunications, and utilities. The data span from January 1995 to August 2016.
- The multi-assets dataset consists of asset classes exhibiting different risk–return characteristics (in local currencies): S&P 500 (U.S. large cap), Russell 2000 (U.S. small-cap), Euro Stoxx 50 (EA large cap), Euro Stoxx Small Cap (EA small-cap), FTSE 100 (U.K. large cap), FTSE Small Cap (U.K. small-cap), France 2-year bonds, France 5-year bonds, France 10-year bonds, France 30-year bonds, U.S. 2-year bonds, U.S. 5-year bonds, U.S. 10-year bonds, U.S. 30-year bonds, MSCI Emerging Markets (dollars), and gold (dollars).
- We chose France over Germany for data availability reasons. A difficult decision was made for fixed-income indexes: coupons are not reinvested, because rates are low and are expected to stay low for a long time. This implies that performances in the future will not come from coupons. Because our aim is to build portfolios that *will* perform and not ones that *have* performed, we prefer this solution. As a consequence, no dividends are reinvested. The data span from February 1989 to August 2016.
- Individual stocks with a sufficiently long historical data from the current S&P 500 compose the last dataset. That gives us 357 series to work with. The objective is to get "real" correlations between stocks. Obviously, this dataset does not incorporate information on delistings. Because there is a strong survivor bias, comparisons with the S&P 500 are meaningless. Nevertheless, comparisons between different models are meaningful. The data span from January 1996 to August 2016.

Although more data history would have been desirable, the different periods cover a number of different market regimes and shocks to the financial markets and the world economy, including the "dot-com" bubble the Great Recession, and the 1994 and 1998 bond market crashes considered in the multi-asset dataset.

Comparison Measures

Given the time series of daily out-of-sample returns generated by each strategy in each dataset, several comparison criteria are computed:

- The *adjusted sharpe ratio* (*ASR*) (Pezier and White [2008])[6] explicitly adjusts for skewness and kurtosis by incorporating a penalty factor for negative skewness and excess kurtosis:

$$ASR = SR\left[1+\left(\frac{\mu_3}{6}\right)SR - \frac{(\mu_4 - 3)}{24}SR^2\right] \tag{16}$$

  where μ_3 and μ_4 are the skewness, and kurtosis of the returns distribution and *SR* denotes the traditional Sharpe ratio ($SR = \frac{\mu - r_f}{\sigma}$, where r_f is the risk-free rate[7]).
- The certainty-equivalent return (*CEQ*) is the risk-free rate of return that the investor is willing to accept instead of undertaking the risky portfolio strategy. DeMiguel et al. [2009] define the *CEQ* as

$$CEQ = (\mu - r_f) - \frac{\gamma}{2}\sigma^2 \tag{17}$$

  where γ is the risk aversion. Results are reported for the case of $\gamma = 1$, but other values of the coefficient of risk aversion are also considered as a robustness check. More precisely, the employed definition of *CEQ* captures the level of expected utility of a mean–variance investor, which is approximately equal to the certainty-equivalent return for an investor with quadratic utility (DeMiguel et al. [2009]). It is the most important number to consider for building profitable portfolios (Levy [2016]).
- The *max drawdown* (*MDD*) is an indicator of permanent loss of capital. It measures the largest single drop from peak to bottom in the value of a portfolio. In brief, the *MDD* offers investors a worst-case scenario.

- The average turnover per rebalancing (TO) is

$$TO = \frac{1}{F}\sum_{t=2}^{F} |w_{i,t} - w_{i,t-1}| \tag{18}$$

where F is the number of out-of-sample forecasts.
- The sum of squared portfolio weights ($SSPW$) used in Goetzmann and Kumar [2008] exhibits the underlying level of diversification in a portfolio and is defined as follows:

$$SSPW = \frac{1}{F}\sum_{t=2}^{F}\sum_{i=1}^{N} w_{i,t}^2 \tag{19}$$

$SSPW$ ranges from 0 to 1, where 1 represents the most concentrated portfolio.

No transaction costs or economic costs generated by the turnover are reported. Indeed, the study of transaction costs is difficult because investors face different fees and the same strategy can be implemented via futures or exchange-traded funds (ETFs), or contracts for difference (CFDs), or cash. Moreover, taxes and the chosen rebalancing strategy influence costs. Nevertheless, high average turnover per rebalancing leads to expensive strategies.

Data Snooping

Data snooping occurs when the same dataset is employed more than once for inference and model selection. It leads to the possibility that any successful results may be spurious because they could be due to chance (White [2000]). In other words, looking long enough and hard enough at a given dataset will often reveal one or more forecasting models that look good but are in fact useless.

To avoid data snooping (White [2000]), we compute the model confidence set (MCS) procedure proposed by Hansen et al. [2011]. The MCS procedure is a model selection algorithm that filters a set of models from a given entirety of models. The resulting set contains the best models with a probability that is no less than $1 - \alpha$, with α being the size of the test (see the appendix for a formal description).

An advantage of the test is that it does not necessarily select a single model; it instead acknowledges possible limitations in the data because the number of models in the set containing the best model will depend on how informative the data are. The MCS aims thus at finding the best model and all models that are indistinguishable from the best.

EMPIRICAL RESULTS

S&P Sectors

Exhibit 2 highlights the attractiveness of hierarchical-clustering-based portfolios, especially the DBHT-based model. It is the only model selected in the best models set $\hat{M}^*_{70\%}$ for both ASR and CEQ. This portfolio is diversified ($SSPW = 0.122$), but the average turnover per rebalancing is elevated in comparison with other models ($TO = 3.45\%$).

The MV is included in $\hat{M}^*_{ASR-20\%}$ but its diversification ratio $SSPW$ is by far the highest of all models: The portfolio is concentrated instead of being diversified.

EXHIBIT 2
Investment Strategies Comparison: *S&P 500 Sectors* (January 1996–August 2016)

	ASR	*CEQ*	*MDD*	*TO*	*SSPW*
EW	0.422	6.81	54.2	–	0.100
MV	0.448[a]	5.75	37.8	1.780	0.480
MDP	0.344	5.28	57.6	1.830	0.217
ERC	0.442	6.40	51.6	0.294	0.107
IVRB	0.428	6.51	49.0	0.474	0.121
SL	0.418	6.53	53.4	0.669	0.115
CL	0.430	6.87[a]	51.0	0.817	0.114
AL	0.421	6.62	52.6	0.762	0.114
WM	0.415	6.49	52.9	0.883	0.149
DBHT	0.449[b]	7.43[b]	59.0	3.450	0.122

Notes: This exhibit reports comparison criteria used to evaluate the quality of the models: the adjusted Sharpe ratio (ASR), the certainty-equivalent return (CEQ) in percent, the max drawdown (MDD) in percent, the average turnover per rebalancing (TO) in percent, and the sum of squared portfolio weights (SSPW). EW is the equal-weight allocation, MV is the minimum-variance allocation, MDP is the most diversified portfolio allocation, ERC is the equal-risk-contribution allocation, IVRB is the inverse-volatility risk budget allocation, SL is the simple-linkage-based allocation, CL is the complete-linkage-based allocation, AL is the average-linkage-based allocation, WM is the Ward's-method-based allocation, DBHT is the directed bubble hierarchical tree-based allocation.
[a] and [b] indicate the model is in the set of best models $\hat{M}^*_{20\%}$ and $\hat{M}^*_{70\%}$, respectively.

CL belongs to $\hat{M}^*_{CEQ-20\%}$. The portfolio is diversified ($SSPW = 0.114$), and the turnover is quite low ($TO = 0.817\%$).

Multi-Assets Dataset

Exhibit 3 paints a contrasting picture: Risk-based portfolios achieve impressive *ASR* along with low *CEQ*. For instance, IVRB constitutes the best models set $\hat{M}^*_{ASR-70\%}$. To highlight the interest of the MCS, it is important to note that IVRB does not obtain the higher *ASR*, yet it is the best model. Moreover, MDP and ERC are selected in $\hat{M}^*_{ASR-20\%}$. That said, risk-based portfolios attain very low *CEQ*, especially IVRB ($CEQ = 0.951$). Above all, they do not produce diversified portfolios. This implies that portfolios are invested almost solely in bonds, thereby being very exposed to shocks from this asset class. This is not the aim of diversified portfolios.

Hierarchical clustering-based portfolios do not face the same problems. AL, SL, and DBHT compose $\hat{M}^*_{CEQ-70\%}$, while delivering reasonably good *ASR*. All portfolios are diversified, and the average turnover per rebalancing is low for SL and AL. Again, DBHT's average turnover per rebalancing is elevated in comparison with other models.

Individual Stocks

Exhibit 4 illustrates that hierarchical clustering-based portfolios outperform risk-based portfolios. Indeed, DBHT is the only model selected in the best models set $\hat{M}^*_{ASR-70\%}$, and the best models set $\hat{M}^*_{CEQ-70\%}$ is only constituted by one model: AL. Both portfolios are diversified.

The main drawback is the surprising elevated average turnover per rebalancing. This point needs to be further investigated—in particular, the impact of the criteria employed to select the number of clusters and

Exhibit 3

Investment Strategies Comparison: *S&P 500 Sectors*, February 1989–August 2016

	ASR	*CEQ*	*MDD*	*TO*	*SSPW*
EW	0.601[a]	4.02	24.9	–	0.0625
MV	0.611	1.31	7.49	4.220	0.403
MDP	0.717[a]	1.95	7.7	2.920	0.296
ERC	0.707[a]	1.91	9.34	0.509	0.164
IVRB	0.581[b]	0.951	6.85	0.500	0.342
SL	0.597	4.67[b]	31.4	1.280	0.086
CL	0.586	4.51	29.9	1.150	0.085
AL	0.602	4.71[b]	29.7	1.190	0.085
WM	0.583	4.47	29.9	1.250	0.084
DBHT	0.525	4.87[b]	25.4	4.350	0.081

Notes: This exhibit reports comparison criteria used to evaluate the quality of the models: the adjusted Sharpe ratio (ASR), the certainty-equivalent return (CEQ) in percent, the max drawdown (MDD) in percent, the average turnover per rebalancing (TO) in percent, and the sum of squared portfolio weights (SSPW). EW is the equal-weight allocation, MV is the minimum-variance allocation, MDP is the most diversified portfolio allocation, ERC is the equal-risk-contribution allocation, IVRB is the inverse-volatility risk budget allocation, SL is the simple-linkage-based allocation, CL is the complete-linkage-based allocation, AL is the average-linkage-based allocation, WM is the Ward's-method-based allocation, DBHT is the directed bubble hierarchical tree–based allocation.

[a] and [b] indicate the model is in the set of best models $\hat{M}^*_{20\%}$ and $\hat{M}^*_{70\%}$, respectively.

Exhibit 4

Investment Strategies Comparison: *Individual Stocks*, January 1996–August 2016

	ASR	*CEQ*	*MDD*	*TO*	*SSPW*
EW	0.595	13.3	52.2	–	0.0028
MV	0.600	13.0	51.2	0.021	0.0048
MDP	0.658	16.4	45.1	7.210	0.052
ERC	0.570	12.1	49.4	0.790	0.0036
IVRB	0.560	10.8	47.1	0.980	0.0045
SL	0.492	19.2[a]	43.1	32.400	0.0552
CL	0.473	16.5	47.4	33.700	0.0151
AL	0.520	19.5[b]	46.1	33.600	0.041
WM	0.572	14.2	51.2	33.400	0.0051
DBHT	0.591[b]	13.5	47.6	39.400	0.0056

Notes: This exhibit reports comparison criteria used to evaluate the quality of the models: the adjusted Sharpe ratio (ASR), the certainty-equivalent return (CEQ) in percent, the max drawdown (MDD) in percent, the average turnover per rebalancing (TO) in percent, and the sum of squared portfolio weights (SSPW). EW is the equal-weight allocation, MV is the minimum-variance allocation, MDP is the most diversified portfolio allocation, ERC is the equal-risk-contribution allocation, IVRB is the inverse-volatility risk budget allocation, SL is the simple-linkage-based allocation, CL is the complete-linkage-based allocation, AL is the average-linkage-based allocation, WM is the Ward's-method-based allocation, DBHT is the directed bubble hierarchical tree–based allocation.

[a] and [b] indicate the model is in the set of best models $\hat{M}^*_{20\%}$ and $\hat{M}^*_{70\%}$ respectively.

the use of "shrinkage" to improve the estimation of the correlation matrix (see Ledoit and Wolf [2004], Ledoit and Wolf [2014], and Gerber et al. [2015]).

CONCLUSION

Diversification is often spoken of as the only free lunch in investing. Yet, truly diversifying a portfolio is easier said than done. For instance, modern portfolio optimization techniques often fail to outperform a basic equal-weighted allocation (DeMiguel et al. [2009]).

Building upon the fundamental notion of hierarchy (Simon [1962]), López de Prado [2016a] introduces a new portfolio diversification technique called hierarchical risk parity, which uses graph theory and machine learning techniques.

Exploiting the same basic idea in a different way, we propose a hierarchical clustering-based asset allocation. Classical and more modern hierarchical clustering methods are tested, namely, simple linkage, complete linkage, average linkage, Ward's method, and the directed bubble hierarchical tree. Once the assets are hierarchically clustered, a simple and efficient capital allocation within and across clusters of investments at multiple hierarchical levels is computed. The main principle is to find a diversified weighting by distributing capital equally to each cluster hierarchy, so that many correlated assets receive the same total allocation as a single uncorrelated one.

The out-of-sample performances of hierarchical-clustering-based portfolios and more traditional risk-based portfolios are evaluated across three empirical datasets, which differ in terms of number of assets and composition of the universe (S&P sectors, multi-assets, and individual stocks). To prevent strategies that perform by luck from being considered effective, we assess the comparison of profit measures using the bootstrap-based model confidence set procedure (Hansen et al. [2011]).

The empirical results point out that hierarchical-clustering-based portfolios are truly diversified and achieve statistically better risk-adjusted performances, as measured by the the adjusted Sharpe ratio (Pezier and White [2008]) and by the certainty-equivalent return on all datasets. The only exception concerns the multi-assets dataset in which risk-based portfolios produce impressive *ASR* along with ridiculously low *CEQ*. Among clustering methods, there is no clear winner: DBHT-based portfolios attain slightly superior risk-adjusted returns, but AL-based portfolios are clearly more robust.

Last but not least, this article opens the door for further research. Testing other clustering methods and investigating typical machine learning issues, such as the choice of the distance measure and the criteria used to select the number of clusters, come naturally to mind. Above all, improving the estimation of the correlation matrix seems to be the most important priority. Potential improvements may come from the use of "shrinkage."

APPENDIX

MODEL CONFIDENCE SET

Define a set M_0 that contains the set of models under evaluation indexed by $i = 0, \ldots, m_0$. Let $d_{i,j,t}$ denote the loss differential between two models by

$$d_{i,j,t} = L_{i,t} - L_{j,t}, \forall i, j \in M_0 \tag{A-1}$$

L is the loss calculated from some loss function for each evaluation point $t = 1, \ldots, F$. The set of superior models is defined as

$$M* = \{i \in M_0 : E[d_{i,j,t}] \leq 0 \ \forall j \in M_0\} \tag{A-2}$$

The MCS uses a sequential testing procedure to determine $M*$. The null hypothesis being tested is

$$\begin{cases} H_{0,M} : E[d_{i,j,t}] = 0 \ \forall i, j \in M \text{ where } M \text{ is a subset of } M_0 \\ H_{A,M} : E[d_{i,j,t}] \neq 0 \text{ for some } i, j \in M \end{cases} \tag{A-3}$$

When the equivalence test rejects the null hypothesis, at least one model in the set M is considered inferior and the model that contributes the most to the rejection of the null is eliminated from the set M. This procedure is repeated until the null is accepted and the remaining models in M now equal $\hat{M}^*_{1-\alpha}$.

According to Hansen et al. [2011], the following two statistics can be used for the sequential testing of the null hypothesis:

$$t_{i,j} = \frac{\bar{d}_{i,j}}{\sqrt{\widehat{var}(\bar{d}_{i,j})}} \text{ and } t_i = \frac{\bar{d}_i}{\sqrt{\widehat{var}(\bar{d}_i)}} \tag{A-4}$$

where m is the number of models in M, $\bar{d}_i = (m-1)^{-1} \Sigma_{j \in M} \bar{d}_{i,j}$, is the simple loss of the *ith* model relative to the averages losses across models in the set M, and $\bar{d}_{i,j} = (m)^{-1} \Sigma_{t=1}^{m} d_{i,j,t}$ measures the relative sample loss between the *ith* and *jth* models. Because the distribution of the test statistic depends on unknown parameters, a bootstrap procedure is used to estimate the distribution.

ENDNOTES

[1]Since the seminal work of Mantegna [1999], correlation networks have been extensively used in econophysics as tools to filter, visualize, and analyze financial market data.

[2]The results of applying K-means or K-medoids clustering algorithms depend on the choice for the number of clusters to be searched and a starting configuration assignment. In contrast, hierarchical clustering methods do not require such specifications.

[3]We consider five if the equal-weighted portfolio is seen as a risk-budgeting portfolio.

[4]For investors, the choice of the rebalancing strategy is crucial. The periodic rebalancing is not optimal, and other options should be investigated (Sun et al. [2006]).

[5]Data are available from the author upon request.

[6]Similar to the adjusted Sharpe ratio, the modified Sharpe ratio uses modified VaR adjusted for skewness and kurtosis as a risk measure.

[7]A risk-free interest rate of zero is assumed when calculating the *ASR* and *CEQ*.

REFERENCES

Alipour, E., C. Adolphs, A. Zaribafiyan, and M. Rounds. "Quantum-Inspired Hierarchical Risk Parity." Unpublished paper, 2016.

Anderson, P.W. "More Is Different." *Science*, Vol. 177, No. 4047 (1972), pp. 393-396.

Aste, T., W. Shaw, and T.D. Matteo. "Correlation Structure and Dynamics in Volatile Markets." *New Journal of Physics*, Vol. 12, No. 8 (2010), pp. 5-9.

Choueifaty, Y., T. Froidure, and J. Reynier. "Properties of the Most Diversified Portfolio." *Journal of Investment Strategies*, Vol. 2, No. 2 (2013), pp. 44-70.

DeMiguel, V., L. Garlappi, and R. Uppal. "Optimal Versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy?" *The Review of Financial Studies*, Vol. 22, No. 5 (2009), pp. 1915-1953.

Gerber, S., H. Markowitz, and P. Pujara. "Enhancing Multi-Asset Portfolio Construction Under Modern Portfolio Theory with a Robust Co-Movement Measure." Unpublished paper, 2015.

Goetzmann, W.N., and A. Kumar. "Equity Portfolio Diversification." *Review of Finance*, Vol. 12, No. 3 (2008), pp. 433-463.

Griveau-Billion, T., J.C. Richard, and T. Roncalli. "A Fast Algorithm for Computing High-Dimensional Risk Parity Portfolios." Unpublished paper, 2013.

Hansen, P., A. Lunde, and J. Nason. "The Model Confidence Set." *Econometrica*, Vol. 79, No. 2 (2011), pp. 453-497.

Lau, A., M. Kolanovic, T. Lee, and R. Krishnamachari. "Cross Asset Portfolios of Tradable Risk Premia Indices." Global Quantitative and Derivatives Strategy, JP Morgan, 2017.

Ledoit, O., and M. Wolf. "Honey, I Shrunk the Sample Covariance Matrix." *The Journal of Portfolio Management*, Vol. 30, No. 4 (2004), pp. 110-119.

——. "Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks." Unpublished paper, 2014.

Levy, M. "Measuring Portfolio Performance: Sharpe, Alpha, or the Geometric Mean?" Unpublished paper, 2016.

López de Prado, M. "Building Diversified Portfolios That Outperform Out of Sample." *The Journal of Portfolio Management*, Vol. 42, No. 4 (2016a), pp. 59-69.

——. Mathematics and Economics: A Reality Check." *The Journal of Portfolio Management*, Vol. 43, No. 1 (2016b), pp. 5-8.

Maillard, S., T. Roncalli, and J. Teiletche. "The Properties of Equally Weighted Risk Contribution Portfolios." *The Journal of Portfolio Management*, Vol. 36, No. 4 (2010), pp. 60-70.

Mantegna, R.N. "Hierarchical Structure in Financial Markets." *The European Physical Journal B: Condensed Matter and Complex Systems*, Vol. 11, No. 1 (1999), pp. 193-197.

Markowitz, H. "Portfolio Selection." *The Journal of Finance*, Vol. 7, No. 1 (1952), pp. 77-91.

Michaud, R. "The Markowitz Optimization Enigma: Is "Optimized" Optimal?" *Financial Analysts Journal*, Vol. 45, No. 1 (1989), pp. 31-42.

Musmeci, N., T. Aste, and T. Di Matteo. "Relation between Financial Market Structure and the Real Economy: Comparison Between Clustering Methods." *PLoS One*, Vol. 10, No. 3 (2015), pp. 201-210.

Pezier, J., and A. White. "The Relative Merits of Alternative Investments in Passive Portfolios." *The Journal of Alternative Investments*, Vol. 10, No. 4 (2008), pp. 37-39.

Roncalli, T. *Introduction to Risk Parity and Budgeting.* Boca Raton, FL: Chapman & Hall, 2013.

Simon, H.A. "The Architecture of Complexity." *Proceedings of the American Philosophical Society*, Vol. 106, No. 6 (1962), pp. 467-482.

Song, W.M., T. Di Matteo, and T. Aste. "Hierarchical Information Clustering by Means of Topologically Embedded Graphs." *PLoS One*, Vol. 7, No. 3 (2012), pp. 41-50.

Sun, W., A. Fan, L.W. Chen, T. Schouwenaars, and M. Albota. "Optimal Rebalancing for Institutional Portfolios." *The Journal of Portfolio Management*, Vol. 32, No. 2 (2006), pp. 33-43.

Tibshirani, R., G. Walther, and T. Hastie. "Estimating the Number of Clusters in a Data Set via the Gap Statistic." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 63, No. 2 (2001), pp. 411-423.

Tumminello, M., T. Aste, T. Di Matteo, and R. Mantegna. "A Tool for Filtering Information in Complex Systems." *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 102, No. 30 (2005), pp. 11-23.

Tumminello, M., F. Lillo, and R. Mantegna. "Correlation, Hierarchies, and Networks in Financial Markets." *Journal of Economic Behavior & Organization*, Vol. 75, No. 1 (2010), pp. 40-58.

Ward, J.H. "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association*, Vol. 58, No. 301 (1963), pp. 236-244.

White, H. "A Reality Check for Data Snooping." *Econometrica*, Vol. 68, No. 5 (2000), pp. 1097-1126.