The Hierarchical Equal Risk Contribution Portfolio

Thomas Raffinot

August 2018

Abstract

Building upon the fundamental notion of hierarchy, the "Hierarchical Risk Parity" (HRP) and the "Hierarchical Clustering based Asset Allocation" (HCAA), the Hierarchical Equal Risk Contribution Portfolio (HERC) aims at diversifying capital allocation and risk allocation. HERC merges and enhances the machine learning approach of HCAA and the Top-Down recursive bisection of HRP. In more detail, the modified Top-Down recursive division is based on the shape of dendrogram, follows an Equal Risk Contribution allocation and is extended to downside risk measures such as conditional value at risk (CVaR) and Conditional Drawdown at Risk (CDaR). The out-of-sample performances of hierarchical clustering based portfolios are evaluated across two empirical datasets, which differ in terms of number of assets and composition of the universe (multi-assets and individual stocks). Empirical results highlight that HERC Portfolios based on downside risk measures achieve statistically better risk-adjusted performances, especially those based on the CDaR.

JEL classifications: G00, G10, G11

Keywords: Hierarchical Clustering; Asset Allocation; Model Confidence Set; Portfolio Construction; Graph Theory; Financial Networks; Machine Learning; Equal Risk Contribution

---

Thomas Raffinot is the Head of Quantitative Macro Strategy at Silex Investment Partners (email: traffinot@gmail.com).

I thank Theophile Charmont for statistical assistance. The views expressed in this paper are the sole responsibility of the authors. Any remaining errors or shortcomings are those of the authors.

Electronic copy available at: https://ssrn.com/abstract=3237540

Nobel Prize winner Herbert Simon argues that complex systems, such as financial markets, have a structure and are usually organized in a hierarchical manner, with separate and separable sub-structures (Simon [1962]). The hierarchical structure of interactions among elements strongly affects the dynamics of complex systems.

However, correlation matrices lack the notion of hierarchy, which allows weights to vary freely in unintended ways (López de Prado [2016]). This is one of the main reasons why modern portfolio optimization techniques often fail to outperform a basic equal-weighted allocation (DeMiguel et al. [2009]).

To alleviate this issue, López de Prado [2016] introduces a portfolio diversification technique called "Hierarchical Risk Parity" (HRP), which uses graph theory and machine learning techniques. Exploiting the same basic idea in a different way, Raffinot [2017a] proposes a hierarchical clustering based asset allocation (HCAA).

Briefly, the principle is to retain the correlations that really matter and once the assets are hierarchically clustered, a capital allocation is estimated. Yet, the design and the implementation of the portfolio optimization differ substantially between the two approaches.

HRP starts by reorganizing the covariance matrix to place similar investments together. Then, it employs an inverse-variance weighting allocation based on the number of assets with no further use of the clustering.

HCAA allocates capital within and across clusters of assets at multiple hierarchical levels. The "Hierarchical 1/N" proposed in Raffinot [2017a] finds a diversified weighting by distributing capital equally among each cluster hierarchy. Then, within a cluster, an equal-weighted allocation is computed. This

Electronic copy available at: https://ssrn.com/abstract=3237540

naive approach inspired by DeMiguel et al. [2009] requires neither expected returns nor risk measures. Yet, it may be too simple since risk management is not part of the weighting strategy.

The objective of the paper is to enhance HCAA by taking the best of both methods. The Hierarchical Equal Risk Contribution Portfolio (HERC) merges and enhances the machine learning approach of HCAA and the Top-Down recursive bisection of HRP. Furthermore, HERC is extended to downside risk measures such as conditional value at risk (CVaR) and Conditional Drawdown at Risk (CDaR).

The performances of hierarchical clustering based portfolios are evaluated across two empirical datasets, which differ in terms of number of assets and composition of the universe (multi-assets and individual stocks).

To avoid data snooping and to develop confidence about profit measures, including the potential for drawdowns and fat tails, synthetic performance histories are simulated with a process called block bootstrapping. The idea is to simulate different paths that the results might have taken, if the ordering had been different.

The findings of the paper can be summarized as follows: the "Hierarchical 1/N" is difficult to beat, but Hierarchical Equal Risk Contribution portfolios based on downside risk measures achieve statistically better risk-adjusted performances, especially those based on the Conditional Drawdown at Risk.

The rest of the paper proceeds as follows. Section 1 introduces the Hierarchical Equal Risk Contribution Portfolio. To this aim, the methods proposed by López de Prado [2016] and by Raffinot [2017a] are reviewed and discussed. Section 2 presents the empirical set up: the datasets, the pitfall of

Electronic copy available at: https://ssrn.com/abstract=3237540

data snooping and the comparison criteria. Section 3 analyses the empirical results.

The Hierarchical Equal Risk Contribution Portfolio

The Hierarchical Clustering based Asset Allocation

The objective of "Hierarchical Clustering based Asset Allocation" (HCAA) is to find a trade-off between diversification across all investments and diversification across clusters of investments at multiple hierarchical levels. The HCAA is computed in four steps:

- Step 1: Hierarchical clustering. Many options exist, especially Simple Linkage, Complete Linkage, Average Linkage and Ward's Method (see Appendix A for a formal description)
- Step 2: Selection of the optimal number of clusters based on the Gap index (Tibshirani et al. [2001])
- Step 3: Capital allocation across clusters
- Step 4: Capital allocation within clusters

The "Hierarchical 1/N" proposed in Raffinot [2017a] attempts to stay very simple and focuses not only on the clusterings, but on the entire hierarchies associated to those clusterings. The principle is to find a diversified weighting by distributing capital equally among each cluster hierarchy, so that many

Electronic copy available at: https://ssrn.com/abstract=3237540

correlated assets receive the same total allocation as a single uncorrelated one. Then, within a cluster, an equal-weighted allocation is computed.

Exhibit 1 illustrates a small dendrogram with five assets and three clusters. Exhibit 2 exhibits the same dendrogram, but with only two clusters. It highlights the importance of the number of clusters on capital allocation.

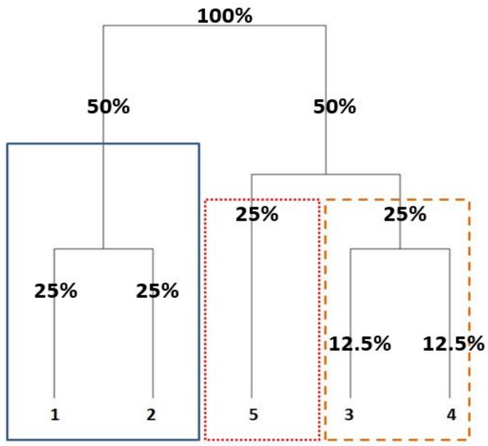

Exhibit 1: Asset allocation weights: a small example

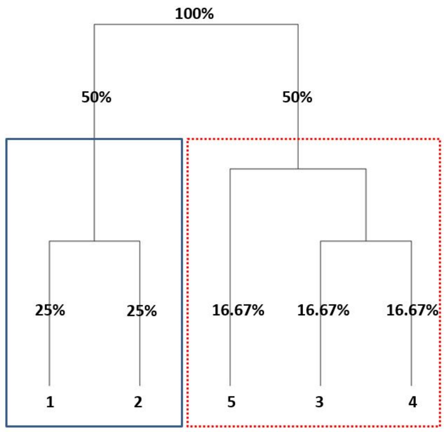

Exhibit 2: Importance of the number of clusters

The chosen weighting scheme does not take risk into account, which can be problematic. Indeed, even if portfolios are diversified in terms of capital allocation, portfolios may be heavily concentrated in terms of risk allocation (Maillard et al. [2010]).

Hierarchical Risk Parity

Hierarchical Risk Parity (HRP) aims at diversifying risk allocation. It operates in three stages:

- Step 1: Minimum Spanning Tree (MST). This procedure allows to

Electronic copy available at: https://ssrn.com/abstract=3237540

extract a MST and a hierarchical tree from a correlation coefficient matrix by means of a well defined algorithm known as Single Linkage clustering algorithm[1]

- Step 2: Quasi-diagonalization: reorganization of the covariance matrix to place similar investments together

- Step 3: Recursive bisection: the matrix diagonalization allows to distribute weights optimally following an inverse-variance allocation between uncorrelated assets. The Top-Down allocation algorithm is as follows:

  - Initialize the weighting to each security to 1, $w_i = 1, i = \{1..n\}$
  - Bisect the portfolio into two sets, s_1 and s_2
  - Let V_i be the covariance matrix for set s_i
  - Let $W_i = diag\left(V_i\right)^{-1} * \frac{1}{tr\left(diag(V_i)^{-1}\right)}$
  - Let $V_{adj,i} = W_i' * V_i * W_i$
  - Let $a_1 = 1 - \frac{V_{adj,1}}{V_{adj,1}+V_{adj,2}}, a_2 = 1 - a_1$
  - Adjust weightings for each set as $w_{s_i} = w_{s_i} * a_i$

The first remark concerns the link between the recursive inverse-variance allocation and the "Equal Risk Contribution" (ERC) portfolio. In the special

[1] A Minimum Spanning Tree is a correlation-network method used in Econophysics as tool to filter, visualise and analyse financial market data. 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. If you have N assets of interest, it keeps $N-1$ links from the $\frac{1}{2}N(N-1)$ pairwise correlations.

Electronic copy available at: https://ssrn.com/abstract=3237540

case of two uncorrelated assets, the minimum-variance portfolio equals the "Equal Risk Contribution" (ERC) portfolio (Maillard et al. [2010]). More generally, in the case of two assets, the ERC solution never depends on the correlation between assets (Maillard et al. [2010]). From a theoretical point of view, it seems more appropriate to make no assumptions on the correlation between clusters, thereby expressing the bisection in terms of ERC rather than in terms of minimum-variance between uncorrelated assets.

The second remark is that HRP stems much more from graph theory than machine learning.

First, HRP is based on the the equivalence between MST and the single-linkage clustering algorithm. But single linkage suffers from chaining. In order to merge two groups, only one pair of points needs to be close, irrespective of all others. Therefore clusters can be too spread out, and not compact enough.

Second, the MST is only used to reorder the assets. The shape of the dendrogram is not considered when the bisection is done. Only the number of assets matters. For example, exhibit 3 illustrates the bisection based on the size of the investment universe and the division into two parts induced by the dendrogram for a simple case composed of four assets and two clusters. The bisection based on the number of assets divides the second cluster in two different parts, which is far from natural from a machine learning point of view.

Electronic copy available at: https://ssrn.com/abstract=3237540

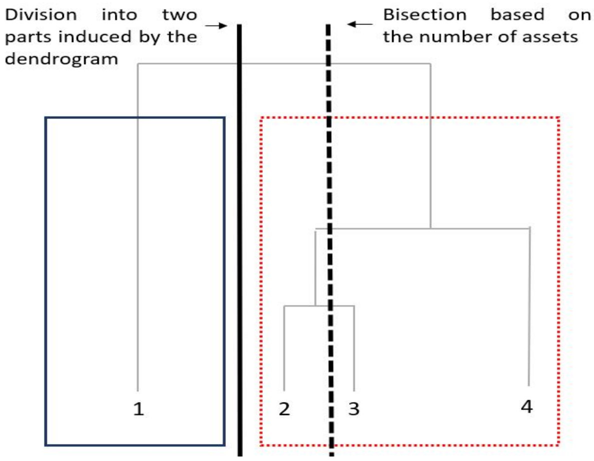

Exhibit 3: Bisection based on the number of assets and bisection induced by the dendrogram

Last, the relevant number of clusters is not considered: each asset is its own cluster. It may happen that you use more information than the clustering can provide with reliability. Intuitively, it can be seen as a form of overfitting, leading to potential bad results. Actually, overfitting denotes the situation when a model targets particular observations rather than a general structure: the model explains the training data instead of finding patterns that could generalize it. In other words, attempting to make the model conform too closely to slightly inaccurate data can infect the model with substantial errors and reduce its predictive power.

The Hierarchical Equal Risk Contribution Portfolio

The Hierarchical Equal Risk Contribution Portfolio (HERC) aims at diversifying capital allocation and risk allocation. It merges the machine learning approach of HCAA and the Top-Down recursive bisection of HRP. It is computed in four stages:

Electronic copy available at: https://ssrn.com/abstract=3237540

- Step 1: Hierarchical clustering

- Step 2: Selection of the optimal number of clusters based on the Gap index (Tibshirani et al. [2001])

- Step 3: Top-Down recursive division into two parts based on the dendrogram and following an Equal Risk Contribution allocation, e.g. the weights are $a_1 = \frac{\mathcal{RC}_1}{\mathcal{RC}_1 + \mathcal{RC}_2}, a_2 = 1 - a_1$ where $\mathcal{RC}_1$ is the risk contribution of the first cluster and $\mathcal{RC}_2$ is the risk contribution of the second cluster

- Step 4: Naive Risk Parity within clusters (within the same cluster the correlation between assets should be elevated)

The variance is the most common risk metric. While variance is a good measure of risk in a world of normally distributed returns, it is not a good measure for the risk of assets with significant tail risk. To alleviate this issue, HERC can easily be extended to downside risk measures such as conditional value at risk (CVaR) and Conditional Drawdown at Risk (CDaR) (refer to Chekhlov et al. [2005] for more information on drawdown measures and asset allocation). The CVaR is derived by taking a weighted average of the losses in the tail of the distribution of possible returns beyond a given quantile, namely the value at risk (VaR). The CDaR is similar to CVaR and can be viewed as a modification of the CVaR to the case when the loss-function is defined as a drawdown. The CDaR and CVaR are conceptually related percentile-based risk performance functionals. Like CVaR, CDaR is defined as the mean of the worst drawdowns above a quantile. In this paper, a

Electronic copy available at: https://ssrn.com/abstract=3237540

quantile of 1% is chosen and risk measures are estimated from past returns.

Investment strategies comparison

Datasets

The performances of models are evaluated across two very disparate datasets. The two considered datasets differ in terms of assets' composition and number of assets[2]:

- The multi-assets dataset is constituted of asset classes exhibiting different risk-return characteristics (in local currencies): S&P 500 (US large cap), Russell 2000 (US small cap), Euro Stoxx 50 (EA large cap), Euro Stoxx Small Cap (EA small cap), FTSE 100 (UK large cap), FTSE Small Cap (UK small cap), France 2-Year bonds, France 5-Year bonds, France 10-Year bonds, France 30-Year bonds, US 2-Year bonds, US 5-Year bonds, US 10-Year bonds, US 30-Year bonds, MSCI Emerging Markets (dollars), Gold (dollars), CRB (dollars).

  As in Raffinot [2017a], coupons and dividends are not reinvested[3]. The data span from February 1989 to June 2018.

- Individual stocks with a sufficiently long historical data from the current S&P 500 compose the second dataset. That gives us 365 series to work with. Since this dataset does not incorporate information on

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

[3]The reasoning is the following: rates are low and are expected to stay low for a long time. It implies that performances in the future will not come from coupons, especially in the euro area.

Electronic copy available at: https://ssrn.com/abstract=3237540

delistings, there is a strong survivor bias. Comparisons with the S&P 500 are thus meaningless. The data span from June 1997 to June 2018.

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, the so-called "taper tantrum" and the 1994 and 1998 bond market crashes as regards the multi-asset dataset.

Data snooping

Data snooping occurs when the same data set 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 data set will often reveal one or more models that look good but are in fact useless.

To alleviate this issue, Raffinot [2017a] employs the model confidence set (MCS) procedure (Hansen et al. [2011]). The MCS aims at finding the best model and all models which are indistinguishable from the best. However, Aparicio and López de Prado [2018] find that MCS requires a very high signal-to-noise ratio to be efficient.

The combinatorial cross-validation (López de Prado [2018]) is another option to limit data snooping. The main advantage of this approach is to be able to get a distribution of the various criteria used to asses the quality of the models.

Briefly, cross-validation uses part of the available data to fit the model, and a different part to test it. K-fold cross-validation works by dividing the

Electronic copy available at: https://ssrn.com/abstract=3237540

training data into K contiguous folds. For the k^{th} part, the learning method is fit to the other $K-1$ parts of the data, and calculate the prediction error of the fitted model when predicting the k^{th} part of the data. This is done for $k = 1, 2, ..., K$ and the K prediction error estimates are averaged.

The combinatorial cross-validation is quite similar to cross-validation, it consists in dividing the dataset in n folds. But, instead of using K-1 folds as the train set, it uses K-t folds. As a result $\binom{K}{t}$ train/test splits can be generated. Each combination involves t tested groups. The total number of the tested groups is thus $t\binom{K}{K-t}$. All combination are computed, which implies that $\frac{t}{K}\binom{K}{K-t}$ complete backtests can be created.

The main drawback is that temporal dynamics are lost. Yet, correlations are time varying and sensitive to economic cycles. In particular, asset classes are more correlated during bad times than during good times. An important implication of higher correlation is that otherwise-diversified portfolios lose some of diversification benefits during bad times, when most needed (see Raffinot [2017b] for more information on this subject).

The cross-validation approach will smooth correlation dynamics. In other words, it will produce portfolios that are very different from what a traditional investor would get in real time.

To get a distribution of the various criteria without loosing the temporal link between assets, a block bootstrap approach has been chosen. The idea is to simulate assets paths by drawing on actual assets paths. It creates thousands of alternative histories to consider how the future may unfold (see Buhlmann [2002]). The block bootstrap is the most general method to improve the accuracy of the bootstrap for time series. The principle is to break

Electronic copy available at: https://ssrn.com/abstract=3237540

the series into roughly equal-length blocks of consecutive observations, to resample the block with replacement, and then to paste the blocks together[4].

For example, if the time series is of length 200 and one uses 10 blocks of length 20, then the blocks are the first 20 observations, the next 20, and so forth. A possible resample is the fourth block (observation 61 to 80), then the last block (observation 181 to 200), then the second block (observation 21 to 40), then the fourth block again, and so on until there are 10 blocks in the resample.

The disadvantage is that past returns represent the full range of possible future returns. Historical simulations are sample dependent and therefore, they are realistic with respect to the sample.

Comparison measures

On each sample, portfolios are updated on a daily basis with no forward-looking biases. The critical estimation window length is a function of the number of assets: 252 for the multi-assets dataset and 504 for the individual stocks dataset. No transactions costs are reported. Indeed, fees, taxes, the implementation of the strategy and the rebalancing policy influence costs. Based on computation time, the number of simulation is set to 500 for the multi-assets dataset and 100 for the individual stocks dataset.

Several comparison criteria are computed:

- The Adjusted Sharpe Ratio ($\mathcal{ASR}$) (Pezier and White [2008]) explicitly adjusts for skewness and kurtosis by incorporating a penalty factor for

[4] It differs from Monte Carlo simulation, since Monte Carlo simulation draws randomly from a theoretical distribution.

Electronic copy available at: https://ssrn.com/abstract=3237540

negative skewness and excess kurtosis:

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

where μ_3 and μ_4 are the skewness and kurtosis of the returns distribution and $\mathcal{SR}$ denotes the traditional Sharpe Ratio ($\mathcal{SR} = \frac{\mu - r_f}{\sigma}$, where r_f is the risk-free rate[5]).

- The certainty-equivalent return ($\mathcal{CEQ}$) 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 $\mathcal{CEQ}$ as:

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

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 $\mathcal{CEQ}$ 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]).

- The Max drawdown ($\mathcal{MDD}$) 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 $\mathcal{MDD}$ offers investors a worst case scenario.

[5] A risk-free interest rate of zero is assumed when calculating the $\mathcal{ASR}$ and $\mathcal{CEQ}$.

Electronic copy available at: https://ssrn.com/abstract=3237540

- The average turnover per rebalancing ($\mathcal{TO}$):

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

where F is the number of out-of-sample forecasts.

- The Sum of Squared Portfolio Weights ($\mathcal{SSPW}$) used in Goetzmann and Kumar [2008] exhibits the underlying level of diversification in a portfolio and is defined as follows:

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

$\mathcal{SSPW}$ ranges from 0 to 1, where 1 represents the most concentrated portfolio.

Empirical results

Based on the conclusions of López de Prado [2016] and Raffinot [2017a], it is taken for granted that hierarchical clustering based portfolios achieve statistically better risk-adjusted performances than commonly used portfolio optimization techniques. These findings are confirmed by other studies (see, among others, Lau et al. [2017] or Alipour et al. [2016]).

Moreover, for the sake of brevity, empirical results are only reported for portfolios based on the Ward's clustering algorithm, in the exception of HRP portfolios, which are computed from the single linkage algorithm. Indeed,

Electronic copy available at: https://ssrn.com/abstract=3237540

unreported results[6] highlight that HRP performances deteriorate when other clustering algorithms are used and portfolios based on the Ward's clustering algorithm perform slightly better for the other methods. Overall, both the reported and unreported results lead to the same practical implications for investors.

Multi-assets dataset

Exhibit 4 points out that HERC portfolios based on downside risk measures outperform other methods. Indeed, they achieve better $\mathcal{ASR}$ along with impressive $\mathcal{CEQ}$. Portfolios are diversified ($\mathcal{SSPW}$=0.136 for the portfolio based on the CVaR and $\mathcal{SSPW}$=0.141 or the portfolio based on the CDaR)

Moreover, it turns out that the "Hierarchical 1/N" is difficult to surpass. It obtains the second best $\mathcal{ASR}$ (0.497) and the third $\mathcal{CEQ}$ at 3.2%.

The HRP portfolio obtains a descent $\mathcal{ASR}$ (0.365) along with a low $\mathcal{CEQ}$ (0.5%). Above all, HRP does not produce diversified portfolios ($\mathcal{SSPW}$=0.363). The $\mathcal{MDD}$ at 6.7% and the low $\mathcal{TO}$ at 6.6% imply that this portfolio is almost only invested in bonds, thereby being very exposed to shocks from this asset class.

The Hierarchical Equal Risk Contribution Portfolio based on the variance is slightly more diversified than the HRP ($\mathcal{SSPW}$=0.289), which leads to an improvement of the $\mathcal{CEQ}$ (1.6%).

[6] Unreported results are available on demand.

Electronic copy available at: https://ssrn.com/abstract=3237540

Exhibit 4: Investment strategies comparison: multi-assets (June 1997-June 2018)

	$\mathcal{ASR}$	$\mathcal{CEQ}$	$\mathcal{MDD}$	$\mathcal{TO}$	$\mathcal{SSPW}$
HRP	0.365	0.005	0.067	0.066	0.363
H_1/N	0.497	0.032	0.221	0.119	0.119
H_Var	0.343	0.016	0.163	0.173	0.289
H_CVaR	0.498	0.045	0.328	0.094	0.136
H_CDaR	0.476	0.043	0.349	0.099	0.141

Note: This table reports comparison criteria used to evaluate the quality of the models: the Adjusted Sharpe Ratio ($\mathcal{ASR}$), the certainty-equivalent return ($\mathcal{CEQ}$) in percent, the Max drawdown ($\mathcal{MDD}$) in percent, the average turnover per rebalancing (($\mathcal{TO}$) in percent, the Sum of Squared Portfolio Weights ($\mathcal{SSPW}$). HRP refers to the Hierarchical Risk Parity Portfolio, H_1/N refers to the Hierarchical 1/N portfolio, H_Var refers to the Hierarchical Equal Risk Contribution Portfolio based on the variance, H_CVaR refers to the Hierarchical Equal Risk Contribution Portfolio based on the CVaR, H_CDaR refers to the Hierarchical Equal Risk Contribution Portfolio based on the CDaR

Exhibit 5 paints a contrasted picture. The fat-tail distribution of HERC portfolios based on downside risk measures is striking. In comparison, other methods are much more stable. Improving the estimation of the downside risk measures might be a possible line of approach to limit this issue. In particular, risk measures estimated from past returns are a backward-looking, while option-implied risk measures are by construction forward-looking. As a matter of fact, DeMiguel et al. [2013] conclude that using option-implied volatility helps to reduce portfolio volatility.

Electronic copy available at: https://ssrn.com/abstract=3237540

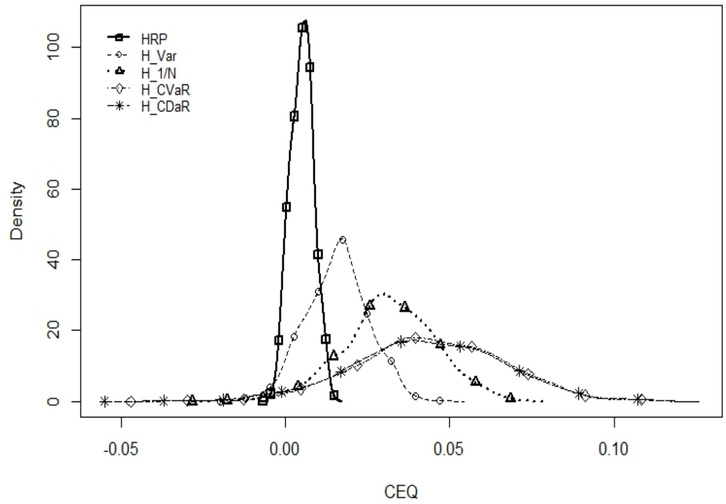

Exhibit 5: Distribution of the certainty-equivalent return for the multi-asset dataset

Individual stocks

Exhibit 6 emphasizes that all methods produce diversified portfolios: the $\mathcal{SSPW}$ criterion lies between 0.004 and 0.045.

Moreover, HERC portfolios based on downside risk measures beat other competitors. Again, they obtain the best $\mathcal{ASR}$ and the best $\mathcal{CEQ}$.

Unlike the previous dataset, HRP performs well with an $\mathcal{ASR}$ of 0.579 and a $\mathcal{CEQ}$ of 11.1%. HRP seems to be much more adapted to the case where assets are homogeneous.

With an $\mathcal{ASR}$ of 0.504 and a $\mathcal{CEQ}$ of 10.8%, the performances of the "Hierarchical 1/N" are close but below those of the best models.

The poor results (in comparison with the others methods) of the HERC portfolio based on the variance are surprising. There is no trivial explanation and this point needs to be further investigated.

Exhibit 6: Investment strategies comparison: individual stocks (February 1989-June 2018)

	$\mathcal{ASR}$	$\mathcal{CEQ}$	$\mathcal{MDD}$	$\mathcal{TO}$	$\mathcal{SSPW}$
HRP	0.579	0.111	0.328	0.223	0.004
H_Var	0.443	0.075	0.448	0.552	0.045
H_1/N	0.504	0.108	0.401	0.445	0.012
H_CVaR	0.599	0.157	0.365	0.332	0.008
H_CDaR	0.594	0.149	0.364	0.328	0.007

Note: This table reports comparison criteria used to evaluate the quality of the models: the Adjusted Sharpe Ratio ($\mathcal{ASR}$), the certainty-equivalent return ($\mathcal{CEQ}$) in percent, the Max drawdown ($\mathcal{MDD}$) in percent, the average turnover per rebalancing (($\mathcal{TO}$) in percent, the Sum of Squared Portfolio Weights ($\mathcal{SSPW}$). HRP refers to the Hierarchical Risk Parity Portfolio, H_1/N refers to the Hierarchical 1/N portfolio, H_Var refers to the Hierarchical Equal Risk Contribution Portfolio based on the variance, H_CVaR refers to the Hierarchical Equal Risk Contribution Portfolio based on the CVaR, H_CDaR refers to the Hierarchical Equal Risk Contribution Portfolio based on the CDaR

Exhibit 7 exhibits that one simulation has been particularly difficult for the HERC portfolio based on the CVaR, which is not the case for the HERC portfolio based on the CDaR.

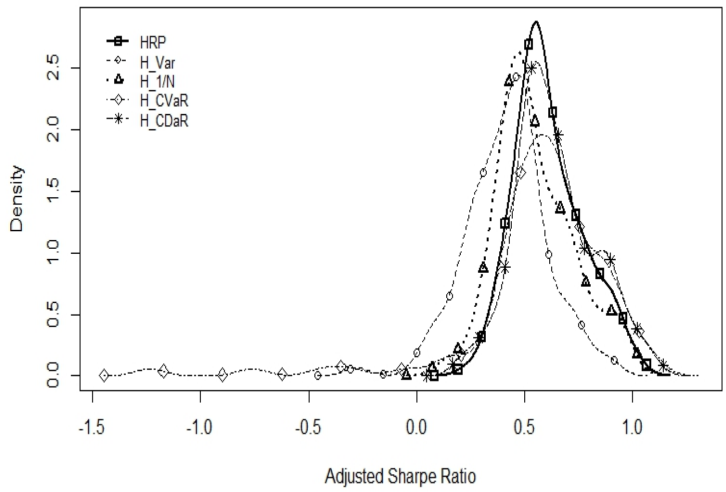

Exhibit 7: Distribution of the adjusted Sharpe Ratio for the individual stocks dataset

Electronic copy available at: https://ssrn.com/abstract=3237540

Conclusion

Building upon the fundamental notion of hierarchy (Simon [1962]), the "Hierarchical Risk Parity" (HRP) (López de Prado [2016]) and the "Hierarchical Clustering based Asset Allocation" (HCAA) (Raffinot [2017a]), the Hierarchical Equal Risk Contribution Portfolio (HERC) aims at diversifying capital allocation and risk allocation.

HERC merges and enhances the machine learning approach of HCAA and the Top-Down recursive bisection of HRP. In more detail, the modified Top-Down recursive division is based on the shape of dendrogram, follows an Equal Risk Contribution allocation and is extended to downside risk measures such as conditional value at risk (CVaR) and Conditional Drawdown at Risk (CDaR).

The out-of-sample performances of hierarchical clustering based portfolios are evaluated across two empirical datasets, which differ in terms of number of assets and composition of the universe (multi-assets and individual stocks). To avoid data snooping and to get a distribution of profit measures, alternative assets paths are simulated with a process called block bootstrapping.

Empirical results highlight that the simple "Hierarchical 1/N" is difficult to beat. Yet, Hierarchical Equal Risk Contribution Portfolios based on downside risk measures achieve statistically better risk-adjusted performances, especially those based on the Conditional Drawdown at Risk.

Last but not least, this article opens the door for further research. Typical machine learning issues have to be investigated, such as the choice of the

Electronic copy available at: https://ssrn.com/abstract=3237540

distance metric and the criteria used to select the number of clusters.

Electronic copy available at: https://ssrn.com/abstract=3237540

Appendix A: Agglomerative clustering

This Appendix from Raffinot [2017a] briefly describes agglomerative 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 requires a suitable distance measure. The following distance is used (Mantegna [1999] and López de Prado [2016]):

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

where $D_{i,j}$ is the correlation-distance index between the i^{th} and j^{th} asset and $\rho_{i,j}$ is the respective Pearson's correlation coefficients. The distance $D_{i,j}$ is a linear multiple of the Euclidean distance between the vectors i, j after z-standardization, hence it inherits the true-metric properties of the Euclidean distance.

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:

Electronic copy available at: https://ssrn.com/abstract=3237540

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

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

  This method is relatively simple and can handle non-elliptical 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 a Minimum Spanning Tree (MST). However the MST retains some information that the SL dendrogram throws away.

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

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

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

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

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

  This is considered to be a fairly robust method.

Electronic copy available at: https://ssrn.com/abstract=3237540

- Ward's Method (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,

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

  where c_i, c_j are the centroids for the clusters.

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

To determine the number of clusters, the Gap index (Tibshirani et al. [2001]) is employed. 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.

References

Alipour, E., Adolphs, C., Zaribafiyan, A., and Rounds, M. (2016). Quantum-inspired hierarchical risk parity. Technical report.

Aparicio, D. and López de Prado, M. (2018). How hard is it to pick the right model? *Algorithmic Finance*.

Buhlmann, P. (2002). Bootstraps for time series. *Statistical Science*, 17(1):52–72.

Chekhlov, A., Uryasev, S., and Zabarankin, M. (2005). Drawdown measure

Electronic copy available at: https://ssrn.com/abstract=3237540

in portfolio optimization. *International Journal of Theoretical and Applied Finance (IJTAF)*, 08(01):13–58.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953.

DeMiguel, V., Plyakha, Y., Uppal, R., and Vilkov, G. (2013). Improving portfolio selection using option-implied volatility and skewness. *Journal of Financial and Quantitative Analysis*, 48(6):1813–1845.

Goetzmann, W. N. and Kumar, A. (2008). Equity portfolio diversification. *Review of Finance*, 12(3):433–463.

Hansen, P., Lunde, A., and Nason, J. (2011). The model confidence set. *Econometrica*, 79(2):453–497.

Lau, A., Kolanovic, M., Lee, T., and Krishnamachari, R. (2017). Cross asset portfolios of tradable risk premia indices. Technical report, Global Quantitative and Derivatives Strategy, JP Morgan.

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

López de Prado, M. (2018). *Advances in Financial Machine Learning*. John Wiley & Sons.

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

Electronic copy available at: https://ssrn.com/abstract=3237540

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

Pezier, J. and White, A. (2008). The relative merits of alternative investments in passive portfolios. *The Journal of Alternative Investments*.

Raffinot, T. (2017a). Hierarchical clustering-based asset allocation. *The Journal of Portfolio Management*, 44(2):89–99.

Raffinot, T. (2017b). Time-varying risk premiums and economic cycles. Technical report.

Simon, H. A. (1962). The architecture of complexity. In *Proceedings of the American Philosophical Society*, pages 467–482.

Tibshirani, R., Walther, G., and Hastie, T. (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.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.

White, H. (2000). A Reality Check for Data Snooping. *Econometrica*, 68(5):1097–1126.

Electronic copy available at: https://ssrn.com/abstract=3237540