The Journal of Finance and Data Science 9 (2023) 100097

Contents lists available at ScienceDirect

The Journal of Finance and Data Science

journal homepage: www.keaipublishing.com/en/journals/jfds

Research article

Portfolio optimization using cellwise robust association measures and clustering methods with application to highly volatile markets

Emmanuel Jordy Menvouta [a], Sven Serneels [b,c], Tim Verdonck [a,d,*]

[a] KU Leuven, Department of Mathematics, Section of Statistics and Data Science, Leuven, Belgium
[b] Gallop Data, Inc., Denver, CO, USA
[c] University of Antwerp, Department of Mathematics, Section of Applied Mathematics, Antwerp, Belgium
[d] University of Antwerp - imec, Department of Mathematics, Section of Applied Mathematics, Antwerp, Belgium

ARTICLE INFO

Keywords:
Portfolio allocation
Risk contribution
Hierarchical clustering
Linkage
Cellwise robustness
Distance correlation

ABSTRACT

This paper introduces the minCluster portfolio, which is a portfolio optimization method combining the optimization of downside risk measures, hierarchical clustering and cellwise robustness. Using cellwise robust association measures, the minCluster portfolio is able to retrieve the underlying hierarchical structure in the data. Furthermore, it provides downside protection by using tail risk measures for portfolio optimization. We show through simulation studies and a real data example that the minCluster portfolio produces better out-of-sample results than mean-variances or other hierarchical clustering based approaches. Cellwise outlier robustness makes the minCluster method particularly suitable for stable optimization of portfolios in highly volatile markets, such as portfolios containing cryptocurrencies.

1. Introduction

Since the seminal paper from Markowitz (1952), investment professionals have been developing methods to build optimal portfolios in a mean-variance setting. Portfolio optimization methods typically rely upon a normality assumption, as well as on the mean return and covariance matrix of the assets. However, as noted by DeMiguel et al. (2009), the estimation error in the portfolio mean return creates instability in the portfolios, which can be exacerbated by assets with non-Gaussian returns that cause high model uncertainty. An example of such assets are cryptocurrencies (CCs), which tend to exhibit very high realized returns, high volatilities and produce a loss of value more often than a gain, but the gains tend to be much higher than the losses (Elendner et al., 2018). Furthermore, the crypto-currency market tends to produce price bubbles (Hafner, 2018) and the price formation is typically driven by human herding behavior (Petukhina et al., 2021). Hence, investors wanting to include them in their portfolios need to apply statistical techniques that can deal with those peculiar properties to accurately manage risk.

Naïve, or *equal weight*, diversification has been found to outperform mean-variance optimization in terms of risk-adjusted returns for CC-only portfolios (Platanakis et al., 2018; Brauneis and Mestel, 2019; Liu, 2019). However, as noted by DeMiguel et al. (2009), this is not

The authors express their gratitude to the esteemed reviewers and editors whose insightful comments and suggestions have played an instrumental role in elevating the quality of this article. Their expertise and conscientious efforts have contributed immensely towards refining the clarity, rigor and overall impact of this work, and we are thankful for their invaluable contributions.

* Corresponding author. University of Antwerp - imec, Department of Mathematics, Section of Applied Mathematics, Antwerp, Belgium.
E-mail address: Tim.Verdonck@uantwerpen.be (T. Verdonck).

Peer review under responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.jfds.2023.100097
Received 24 November 2022; Received in revised form 4 March 2023; Accepted 12 April 2023
Available online 19 April 2023

surprising given the high model uncertainty and low amount of historical data available for the cryptocurrency market. Hence, different portfolio optimization methods that do not explicitly rely on the estimation of the expected return vector need to be used for CCs.

A recent trend in portfolio optimization is the use of reinforcement learning for portfolio optimization (Yu et al., 2019; Benhamou et al., 2021; Chaouki et al., 2020). These methods represent the portfolio optimisation problem as a stochastic control problem, where the goal is to train an agent which chooses its action to maximise expected accumulated rewards. They are flexible, can accommodate various types of constraints, consider multiple objectives and integrate additional sources of information for the portfolio selection. They also tend to outperform in dynamic environment when market conditions are changing. However, these methods are complex, require vast amount of data to be able to train accurately, and suffer from a lack of explainability as it is not explicitly clear how the portfolio weights are selected. Another branch of machine learning for portfolio optimization is the use of supervised (Pinelis and Ruppert, 2022) and unsupervised machine learning techniques (López de Prado, 2016; Raffinot, 2017; Sass and Thös, 2021), to obtain portfolios that outperform naïve or mean-variance approaches out of sample. The strength of these techniques is that they do not rely on the estimation of the expected return or the inversion of an ill-conditioned covariance matrix and can exploit the hierarchical structure between asset returns. The purpose of the hierarchical structure is to demonstrate how investors construct diversified portfolios in real-life situations. Assets are classified into groups based on factors such as liquidity, industry, size, or region, ensuring that assets within each group are interchangeable. Capital allocation is carried out within each group. In contrast to using external factors, hierarchical clustering creates the grouping based on the asset's returns. This requires less work when grouping the assets and can discover relationships between assets that are based on additional factors. Hierarchical Risk Parity (López de Prado, 2016) was applied to CC-only portfolios and found to outperform mean-variance portfolios (Burggraf, 2021). However, the previously cited machine learning techniques rely upon the Pearson correlation coefficient to estimate similarities between assets, which does not capture non-linear relationships between assets and is known to be sensible to outliers.

As noted by Elendner et al. (2018), CCs frequently produce very large returns of either sign. These outlying events may impact the returns due to one-off events, but may not impact the correlation structure between the different assets (Boudt et al., 2013). This will then lead to estimation errors in the machine learning portfolio optimization methods that rely upon a correct association structure to distribute the weights to the assets. Hence, we need to build models that accurately model the associations between assets in the presence of frequent outlying events. Such robust statistical techniques have been applied for portfolio optimization under i.i.d assumptions in the generalized hyperbolic framework (Hellmich and Kassberger, 2011) and the mixed normal framework (Gambacciani and Paolella, 2017). However, CCs have a strong deviation from normality and other parametric distributions. Hence, our first aim is to propose cellwise robust association measures that are valid under any distribution. Our second aim is to propose a portfolio optimization method that uses such a robust association measure and produces out-of-sample performance significantly better than the equal weight portfolio as in Sass and Thös (2021).

Our contributions in this paper are twofold: at first, we propose the use of cellwise robust association measures for hierarchical clustering. Secondly, we develop a portfolio optimization method that combines cellwise robust association measures, hierarchical clustering and downside risk measures. We believe that we have introduced the first portfolio optimization method that uses a general association measure between assets and adjusts for outliers in the asset returns data which can appear independently. This adjustment is aimed at improving out-of-sample performance. Additionally, this is the first study to use both robust statistical methods and interpretable machine learning techniques in the field of portfolio optimization.

The article is organised as follows: Section 2 introduces the association measures, cellwise contamination model and the transformation for cellwise robustness. Section 3 explains the use of cellwise robust association measures for hierarchical clustering. Section 4 introduces the different asset allocation models under consideration while Section 5 introduces the performance measures that will be used to assess the portfolios. In Section 6, we show through simulation studies, the attractive properties of the proposed method. Section 7 illustrates a real data example and Section 8 concludes.

2. Robust measures of association

In this section, we present association measures that will be used in later sections to build optimal portfolios. Let $Z = (Z_1, \ldots, Z_N) \in \mathbb{R}^N$ be the random variable representing asset returns and let $\mathbf{Z} \in \mathbb{R}^{P \times N}$ denote the matrix containing realisations of the log returns of the N assets over a period of P days. Now the association measures will be applied to the columns of $\mathbf{Z}$, such that they represent the association between components of Z.

2.1. Covariance and correlation

The most widely used measures of association are the product moment (PM) covariance and correlation. The Pearson covariance and correlations between two components Z_l and Z_k of Z, are given by

$$\mathrm{Cov}(Z_l, Z_k) = \mathbb{E}[(Z_l - \mathbb{E}[Z_l])(Z_k - \mathbb{E}[Z_k])] \tag{1}$$

$$\mathrm{Cor}\ (Z_l, Z_k) = \frac{\mathrm{Cov}\ (Z_l, Z_k)}{\sqrt{\mathrm{Var}\ [Z_l]\ \mathrm{Var}\ [Z_k]}} \tag{2}$$

It is well known that the product moment correlation has zero breakdown point, meaning that a single outlier can produce a completely erroneous value.

2.2. Distance covariance and distance correlation

Distance covariance (DCOV, Székely and Rizzo (2013)) is a measure of dependence between random variables that is more general than the product moment covariance, as it can detect both linear and non linear associations. Assume that $\mathbb{E}|Z_l|$ and $\mathbb{E}|Z_k|$ are finite, real random variables, with the norm $|.|$ defined for complex valued functions $f(.), as|f| = f\bar{f}$, and $\bar{f}$ is the complex conjugate of f. The squared DCOV between Z_l and Z_k is given by

$$\begin{aligned}\mathfrak{V}^2(Z_l,Z_k) &= \mathbb{E}|Z_l - Z_l'||Z_k - Z_k'| + \mathbb{E}|Z_l - Z_l'|\mathbb{E}|Z_k - Z_k'| \\ &-\mathbb{E}|Z_l - Z_l'||Z_k - Z_k''| - \mathbb{E}|Z_l - Z_l''||Z_k - Z_k'|.\end{aligned} \tag{3}$$

with (Z_l',Z_k'), (Z_l'',Z_k''), being i.i.d. copies of (Z_l, Z_k). Similarly, their distance correlation (DCOR) is a standardised value between [0, 1] and given by

$$\mathfrak{R}(Z_l,Z_k) = \begin{cases}\dfrac{\mathfrak{V}(Z_l,Z_k)}{\sqrt{\mathfrak{V}(Z_l,Z_l)\mathfrak{V}(Z_k,Z_k)}}, & \text{if } \mathfrak{V}(Z_l,Z_l)\mathfrak{V}(Z_k,Z_k) > 0 \\ 0, & \text{otherwise}\end{cases} \tag{4}$$

DCOV and DCOR have the zero equivariance property, meaning that $\mathfrak{V}(Z_l,Z_k) = 0 \Leftrightarrow \mathfrak{R}(Z_l,Z_k) = 0 \Leftrightarrow Z_l \perp\!\!\!\perp Z_k$, where $\perp\!\!\!\perp$ denotes statistical independence. This property makes it an ideal measure to detect linear and nonlinear associations. Given the paired sample $\{(\mathbf{Z}_{l,1},\mathbf{Z}_{k,1}),\dots,\mathbf{Z}_{l,P},\mathbf{Z}_{k,P})\}$, the sample versions of $\mathfrak{V}^2(Z_l,Z_k)$ and $\mathfrak{R}^2(Z_l,Z_k)$ are given by

$$\mathfrak{V}^2(\mathbf{Z}_l,\mathbf{Z}_k) = \frac{1}{P^2}\sum_{u,r=1}^{P}\widehat{A}_{ur}\widehat{B}_{ur}.$$

$$\mathfrak{R}^2(\mathbf{Z}_l,\mathbf{Z}_k) = \begin{cases}\dfrac{\mathfrak{V}^2(\mathbf{Z}_l,\mathbf{Z}_k)}{\sqrt{\mathfrak{V}^2(\mathbf{Z}_l,\mathbf{Z}_l)\mathfrak{V}^2(\mathbf{Z}_k,\mathbf{Z}_k)}}, & \text{if } \mathfrak{V}^2(\mathbf{Z}_l,\mathbf{Z}_l)\mathfrak{V}^2(\mathbf{Z}_k,\mathbf{Z}_k) > 0 \\ 0, & \text{otherwise}\end{cases}$$

where

$$\begin{aligned}\widehat{A}_{u,r} &= a_{ur} - \bar{a}_{u\cdot} - \bar{a}_{\cdot r} + \bar{a}_{\cdot\cdot}, \quad u,r = 1,\dots,P \\ a_{ur} &= |\mathbf{Z}_{l,u} - \mathbf{Z}_{l,r}|; \quad \bar{a}_{u\cdot} = \frac{1}{P}\sum_{r=1}^{P} a_{ur} \\ \bar{a}_{\cdot r} &= \frac{1}{P}\sum_{u=1}^{P} a_{ur}; \quad \bar{a}_{\cdot\cdot} = \frac{1}{P^2}\sum_{u,r=1}^{P} a_{ur}.\end{aligned}$$

Similarly, we have $\widehat{B}_{u,r} = b_{ur} - \bar{b}_{u\cdot} - \bar{b}_{\cdot r} + \bar{b}_{\cdot\cdot}$, with $u,r \in [1,P] \subset \mathbb{N}_0^+$ with $b_{ur} = |\mathbf{Z}_{k,u} - \mathbf{Z}_{k,r}|$. Owing to these properties, DCOV and DCOR can be implemented efficiently. Implementations of both of these association measures and dimension reduction techniques based upon them, such as the ones investigated by Menvouta et al. (2022) can be found in the direpack package Menvouta et al. (2023).

Despite their desirable properties in terms of association measures, DCOV and DCOR do not lead to positive semi definite matrices and hence cannot be used in mean-variance portfolio optimization methods. Similarly to the product moment measures, DCOV and DCOR have zero breakdown point (Raymaekers and Rousseeuw, 2021). Hence, if we want to use these association measures, we need to remove or diminish the influence of cellwise outlying observations.

2.3. Cellwise contamination model

In a financial setting, the cellwise contamination model of Alqallaf et al. (2009) is more appropriate, as outlying returns can affect either or only one asset if the shock is idiosyncratic, or can affect the market as a whole if the shock is due to external news. Hence, it is more accurate to assume that most of the assets have some contamination in their returns. In particular, assume that Z, V and W, are N-dimensional random vectors, with V following some distribution F with mean μ and scatter matrix Σ and W follows an arbitrary distribution from which the outliers are drawn. Then Z follows the model

$$Z = (1 - \mathbf{B})V + \mathbf{B}W \tag{5}$$

where $\mathbf{B} = \text{diag}(B_1, B_2, \dots, B_N)$ is a diagonal matrix consisting of Bernoulli random variables with $\mathbb{P}(B_i = 1) = \varepsilon_i$, the fraction of contamination in the ith column. Similarly to Alqallaf et al. (2009), we assume independence between V, W and $\mathbf{B}$. The casewise contamination model, where all assets are simultaneously affected by outliers, can be obtained as a special case of such a cellwise model by assuming $\mathbb{P}(B_1 = \dots = B_N) = 1$. We also obtain the fully independent model by assuming independence between $B_1, B_2 \dots, B_N$. If we assume that $\mathbb{P}(B_i = 1) = \varepsilon_i = \varepsilon$, for each component then the probability that an observation is uncontaminated in each of its components is $(1-\varepsilon)^N$, such that even for moderate values of N, there might be no uncontaminated observation. Hence, association measures that are casewise robust might have a low breakdown point in such a scenario, requiring the use of cellwise robust association measures, which are robust against cellwise contamination and correctly reflect the association between the components.

2.4. Transformations for cellwise robustness

Our aim is to adopt association measures in portfolio optimization that are robust to outliers, capture nonlinear associations and are fast to compute, given the size of the universe of assets. We follow Raymaekers and Rousseeuw (2021) and use transformations for the association measures, which has the effect of reducing or removing the influence of outlying observations. In particular, if $\{\mathbf{Z}_{l,1}, \ldots, \mathbf{Z}_{l,P}\}$ are the observations of $\mathbf{Z}_l$, we will first transform each element $\mathbf{Z}_{l,u}$ of $\mathbf{Z}_l$ to $g(\mathbf{Z}_{l,u}) = \psi\left((\mathbf{Z}_{l,u} - \hat{\mu}_l)/\hat{\sigma}_l\right)$, where $\hat{\mu}_l$, $\hat{\sigma}_l$ are respectively robust location and scale estimators of Z_l and the choice of ψ will determine the robustness properties of the transformation. We will then apply the association measures on the transformed variables $h(g(\mathbf{Z}_l))$ and $h(g(\mathbf{Z}_k))$, where $h(g(\mathbf{Z}_{l,u})) = \hat{\mu}_l + \hat{\sigma}_l g(\mathbf{Z}_{l,u})$. This transformation projects the variables onto a space where outliers have little or no influence. In what follows, the association will then be estimated from that transformed space. A first transformation is a so-called wrapping transformation, where ψ is chosen to be a redescending function which has the property of removing the influence of outliers far away from the population location. Such redescending weight functions are commonly used for robust M-estimation of location and scale (Maronna et al., 2006) and have also been proven very effective in the context of dimension reduction, for instance in partial robust M regression (Serneels et al., 2005) or its sparse counterpart (Hoffmann et al., 2015). The ψ function is chosen so as to minimize the supremum of the Change of variance curve (CVC) that measures the stability of the estimator's variance with respect to contamination by outliers. In the case of product moment correlation, let the functional $T_\psi = \text{cor}(\psi(Z_l), \psi(Z_k))$, ε the fraction of contamination, Δ the Dirac measure, V the asymptotic variance of the estimator and F a given bivariate distribution, then

$$\text{CVC}(z, T_\psi, F) = \frac{\partial}{\partial \varepsilon}\left[\log V(T_\psi, (1-\varepsilon)F + \varepsilon(\Delta_z + \Delta_{-z})/2)\right]\Big|_{\varepsilon=0} \tag{6}$$

$$\kappa^*(T_\psi) = \sup_z \text{CVC}(z, T_\psi, F) \tag{7}$$

The redescending ψ function with the highest efficiency for a given $\kappa^*(T_\psi)$ in the context of location estimators is presented in Hampel et al. (1981) and is given by

$$\psi_{b,c}(z) = \begin{cases} z & \text{if } 0 \le |z| \le b \\ q_1 \tanh(q_2(c - |z|))\text{sign}(z), & \text{if } b \le |z| \le c \\ 0 & \text{if } c \le |z|. \end{cases} \tag{8}$$

The optimal values of q_1 and q_2 for $0 < b < c$ are derived in Raymaekers and Rousseeuw (2021) and the default choice is set to $b = 1.5$ and $c = 4$ as this choice produces a good balance between robustness and efficiency. Fig. 1 illustrates the wrapping methodology with the chosen psi function in Equation (8), where a majority of the observations, those in [-b,b] are left unchanged, observations outside [-c, c] do not play a role and the influence of observations in between are reduced.

Unfortunately, the ψ function used for product moment correlation cannot be used for DCOR, since it is not strictly monotonous and thus the zero equivariance property on the wrapped variables will not be transferred to the unwrapped variables. Thus we will use as ψ function that is bounded yet strictly monotone. We choose $\psi(z) = \tanh(z)$, the hyperbolic tangent function.

3. Hierarchical clustering

The aim of cluster analysis is to find groups in data such that the elements in the same group have a high degree of similarity whereas elements in distinct groups have a low degree of similarity. Hierarchical clustering methods create a hierarchy of clusters ranging from a

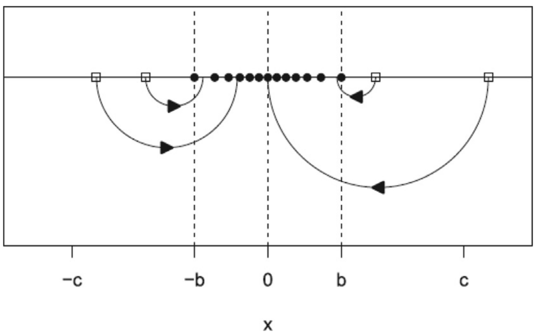

Fig. 1. Illustration of wrapping on standard normal data from Raymaekers and Rousseeuw (2021) with values in [-b,b] not transformed, values outside [-c,c] transformed to zero, and the influence of values in between is reduced.

single cluster containing all observations, to n distinct clusters where each observation is its own cluster. The optimal number of clusters needs to be chosen either by domain knowledge or estimated by using statistical techniques such as the Gap Index (Tibshirani et al., 2001) or the Silhouette plot (Rousseeuw, 1987). For hierarchical clustering, we represent the distance between the ith and jth asset by $D(Z_i, Z_j) = \sqrt{0.5(1-\rho_{i,j})}$ with $\rho_{i,j}$ representing the association measure between Z_i and Z_j. In this paper, we consider agglomerative clustering where each asset starts as its own cluster and the hierarchy of clusters are then progressively formed at each step by merging the two clusters that are the most similar. The measures of dissimilarity between clusters that we will discuss in this article are single linkage, complete linkage and average linkage:

- Single linkage defines the dissimilarity between clusters C_i and C_j by

$$d_{C_i,C_j} = \min_{Z_i,Z_j} \{D(Z_i, Z_j) | Z_i \in C_i, Z_j \in C_j\}.$$

- Complete linkage defines the dissimilarity between clusters C_i and C_j by

$$d_{C_i,C_j} = \max_{Z_i,Z_j} \{D(Z_i, Z_j) | Z_i \in C_i, Z_j \in C_j\}.$$

- Average linkage defines the dissimilarity between clusters C_i and C_j by

$$d_{C_i,C_j} = \underset{Z_i,Z_j}{\text{mean}} \{D(Z_i, Z_j) | Z_i \in C_i, Z_j \in C_j\}.$$

4. Asset allocation models

Let $w = (w_1, \ldots, w_N)^T$ denote the weights of the assets in the portfolio. In what follows, $w \in W := \{w \in \mathbb{R}^N | w_i \geq 0, w^T 1 = 1\}$, which implies that short-selling will not be considered. Note that this constraint is set only to reflect common practical limitations with short-selling cryptocurrency as the methods put forward in this paper are equally suited to a setting that allows for it. Without loss of generality, the risk-free rate will be set to zero. In this section, the different portfolio optimization methods are presented. The different portfolio weights are estimated in-sample the out-of-sample performance of the methods is reported.

4.1. Equally weighted portfolio

The naïve portfolio is given by $w = (\frac{1}{N}, \ldots, \frac{1}{N})^T$. This portfolio requires no estimation or assumptions and distributes the weights equally. It has been found to be optimal out-of-sample in the presence of model uncertainty (Pflug et al., 2012).

4.2. Minimum-variance portfolio

As stated by Markowitz (1952), mean variance portfolios are optimal under a normality assumption for the returns. However, these portfolios are unstable and suffer from low out-of-sample performances due to the estimation error in the mean vector. The minimum variance portfolio, which does not require the estimation of the asset mean return vector, has been shown to outperform mean variance portfolios based on out-of-sample Sharpe ratios (DeMiguel et al., 2009). The weights are obtained by solving the following optimization problem

$$\underset{w}{\text{argmin}} \quad w^T \Sigma w \tag{9}$$

subject to:

$$w^T 1 = 1$$

with Σ representing the covariance matrix of the returns.

4.3. Hierarchical Risk Parity

Hierarchical Risk Parity (HRP) (López de Prado, 2016) is a portfolio optimization method using graph theory and machine learning to allocate weights according to the inverse variance of the assets in a cluster. These clusters are formed using hierarchical clustering on a dissimilarity metric based on Pearson correlation between asset returns. More specifically, the HRP algorithm is as follows:

1. Tree clustering:
   (a) Compute the (distance) correlation matrix $\rho = \{\rho[Z_i, Z_j]\}_{i,j=1,\ldots,N}$.
   (b) Define the dissimilarity matrix $D = \{d_{i,j}\}_{i,j=1,\ldots,N}$ with $d_{i,j} = d(Z_i, Z_j) = \sqrt{0.5(1-\rho_{i,j})}$.

(c) Compute the Euclidean distance between any two columns to construct $\widetilde{D} = \{\widetilde{d}_{i,j}\}_{i,j=1,\dots,N}$ with $\widetilde{d}_{i,j} = \sqrt{\sum_{n=1}^{N}(d_{n,i} - d_{n,j})^2}$

(d) Perform a hierarchical clustering with single linkage on the columns of $\widetilde{D}$.

2. Matrix seriation:At this stage, the covariance matrix of the returns, Σ is rearranged so that similar assets as described by the previous step are closer together compared to dissimilar assets and large covariance elements lie on the diagonal. This has the advantage of providing a quasi-diagonal covariance matrix.
3. Recursive bisection:The previous step produces a quasi-diagonal covariance matrix for which the inverse variance allocation is optimal. In this step, we distribute the weights according to an inverse-variance allocation between adjacent subsets:
   (a) Initialize to a cluster containing all the assets, $C = C_0 = \{i\}_{i=1 \dots, N}$ and initialize weight to $w_i = 1$
   (b) Stop if $|C_n| = 1, \forall C_n \in C$ i.e. if each asset is its own cluster.
   (c) For each $C_n \in C$ such that $|C_n| > 1$, bisect C_n into two clusters C_n^1 and C_n^2.
   (d) For j = 1,2, compute $\widetilde{\Sigma}_n^j = \widetilde{w}_n^{j^T} \Sigma_n^j \widetilde{w}_n^j$ with Σ_i^j the covariance of the elements of C_n^j and $\widetilde{w}_n^j = diag[\Sigma_n^j]^{-1} \frac{1}{tr[diag[\Sigma_n^j]^{-1}]}$
   (e) Compute $\alpha_n^1 = 1 - \frac{\widetilde{\Sigma}_i^1}{\widetilde{\Sigma}_i^1 + \widetilde{\Sigma}_2^j}$; $\alpha_n^2 = 1 - \alpha_n^1$
   (f) Update the weights for elements in C_n^j by multiplying them by α_n^j
   (g) Repeat steps (b) to (g).

The methodology introduced here improves upon HRP by replacing the correlation coefficient in the tree clustering step by the pairwise distance correlation matrix.

4.4. Hierarchical Equal Risk contribution

Raffinot (2018) proposes the Hierarchical Equal Risk contribution (HERC) to overcome some of the shortcomings of HRP portfolios: at first, the inverse variance allocation of HRP is optimal if no correlation between different clusters is assumed. Secondly, HRP uses single linkage clustering, which can suffer from chaining effects whereby clusters get merge because they have two close-by points without taking into account the shape of the clusters or how distant other points in the clusters are. Thirdly, the number of clusters and shape of the dendogram is not considered, which can lead to overfitting and produce unintuitve clusters. The HERC algorithm is as follows:

1. Tree clustering as in HRP. Here, the optimal number of clusters is chosen using the Gap index (Tibshirani et al., 2001).
2. Recursive bisection: This step is similar to that of HRP, however, instead of computing the covariance of each cluster in step 3 (c), we compute each asset's Risk contribution. Hence, steps 3 (d-f) are replaced by:
   (d) For $j = 1, 2$ compute the risk contribution RC_n^j of the elements of C_n^j. The risk measures used for HERC are downside risk measures such as Conditional Value at Risk and Conditional Drawdown at Risk (Chekhlov et al., 2005). Tail risk measures (Chen and Cheng, 2022) can also be used here.
   (e) Compute $\alpha_n^1 = 1 - \frac{RC_n^1}{RC_n^1 + RC_n^2}$ and $\alpha_n^2 = 1 - \alpha_n^1$
   (f) Naïve risk parity: inside each cluster, asset weights are assigned according to the inverse of their risk. That inverse risk weight is then multiplied by the corresponding α_n^j coefficient of the cluster.

4.5. maxCluster portfolio

The maxCluster portfolio was proposed by Sass and Thös (2021) in order to reduce the number of assets used in the Equal weight portfolio. Instead of using the *N* assets for equal weighting, only a subset of the assets is used in the following way:

1. Choose an appropriate number of clusters to be used for clustering.
2. Perform hierarchical clustering with complete linkage of the assets into the chosen number of groups.
3. In each cluster, select as cluster representative the asset with the maximum Sharpe ratio.
4. Perform equal weighting with the cluster representatives.

4.6. minCluster portfolio

The use of the Sharpe ratio to order the assets in each cluster is not optimal as the Sharpe ratio is not a proper ranking measure in a non-gaussian world. Therefore, we propose the minCluster portfolio that uses the Conditional Value at risk, which is a coherent risk measure (Artzner et al., 1999). Moreover, imposing a predetermined number of clusters might not be optimal as we do not know beforehand how many clusters exist in the data. Lastly, the use of complete linkage might produce clusters that are artificially separated and isolated. The proposed strategy for the minCluster portfolio is the following:

1. Use the silhouette plot (Rousseeuw, 1987) to determine the optimal number of clusters.
2. Perform hierarchical clustering with complete linkage of the assets into the optimal number of groups.

3. In each cluster, select as cluster representative the asset with the minimum Conditional-Value-at-Risk at 95%.

We note that the choice of the bottom-up hierarchical clustering is not restrictive, as in our applications, a top-down approach produces similar results.

5. Performance measures

Several performance measures are computed (all assuming a risk free rate r_0 of 0):

- The volatility (standard deviation).
- The Sharpe ratio defined as

$$\mathrm{SR} := \frac{\mu_p - r_0}{\sigma_p}$$

- The adjusted Sharpe ratio (ASR) of Pézier and White (2008) defined as

$$\mathrm{ASR} := \mathrm{SR}\left(1 + \frac{\mathrm{Skew}(R_p)}{6}\mathrm{SR} - \frac{\mathrm{Kurt}(R_p)}{24}\mathrm{SR}^2\right)$$

The traditional Sharpe Ratio may not accurately capture the risk-adjusted returns of investments with non-normal returns distributions, which can be skewed or have heavy tails. The ASR adjusts the Sharpe ratio to account for skewness and kurtosis in the returns distribution, by adding a penalty for negative skewness and excess kurtosis which can significantly impact the Sharpe ratio's accuracy in capturing the risk-adjusted returns.

- The Omega ratio (OR) (Keating and Shadwick, 2002) is a ratio of the portfolio's probability weighted gains relative to its probability weighted losses and is given by

$$\mathrm{OR} := 1 + \frac{\mu_p - r_0}{\mathbb{E}[\max(r_0 - R_p, 0)]}$$

The Omega ratio measures the likelihood of achieving a certain target return by comparing the expected value of the portfolio's positive returns to its expected loss below a specified threshold. The target return is set here to the risk-free rate, and the threshold is a minimum acceptable return set here to zero. As compared to the SR, the OR uses the full return distribution and provides a more comprehensive assessment of the risk-return trade-off by taking into account both the positive and negative returns. A higher Omega ratio means that the expected positive returns are greater than the expected losses, indicating that the portfolio is more likely to achieve the target return.

- The max drawdown (MD) is an indicator of permanent loss of capital. It measures the largest single drop from peak to bottom in the value of a portfolio. MD is represented in %.
- To assess the stability and associated transaction costs of the portfolios, we report the portfolio turnover, defined as

$$\mathrm{TO} := \frac{1}{H-1}\sum_{t=1}^{H-1}\sum_{i=1}^{N}|w_{i,t+1} - w_{i,t+}|$$

where H is the number of rebalancing periods and $w_{i,t+}$ is the weight of asset i in the portfolio before rebalancing at time $t+1$.

- The sum of squared portfolio weights (SSPW) used in Goetzmann and Kumar (2008) is a measure of portfolio diversification ranges between 0 and 1(least diversified) is defined as

$$\mathrm{SSPW} := \frac{1}{H}\sum_{t=1}^{H}\sum_{i=1}^{N}w_{i,t}^2$$

6. Simulation study

6.1. *Robust association measures*

In the first simulation study, we investigate the use of the robust association measures when non-linear associations are present and under cellwise contamination. In particular, we assess how the different measures can recover the correct partitioning of variables using hierarchical clustering methods. Our aim is to see to which extent the wrapping transformation helps capture the true association between variables when cellwise outliers are present. We generate the data as in Kojadinovic (2004), where Z_1 has a uniform

Table 1
Fraction of correctly grouped variables with varying number of observations (n).

Method	n = 50	n = 100	n = 500
Cor	0.69	0.70	0.74
dCor	**1.00**	**1.00**	**1.00**
wCor	0.61	0.66	0.73
wdCor	**1.00**	**1.00**	**1.00**

The bold values represent which method has the best results for the corresponding column.

distribution on $[-1, 1]$, Z_4 and Z_7 are standard normal and all three variables are stochastically mutually independent. The other 6 variables are functions of Z_1, Z_4, and Z_7 as follows:

$$
\begin{aligned}
Z_2 &:= \tanh(Z_1) + Z_1^2 + \varepsilon_2 \\
Z_3 &:= 2\sin(|Z_1|) + \varepsilon_3 \\
Z_5 &:= \sin(Z_4) + \tanh(Z_4) + \varepsilon_5 \\
Z_6 &:= Z_4^2 + \varepsilon_6 \\
Z_8 &:= |Z_7| + \varepsilon_8 \\
Z_9 &:= \sin(|Z_7|) + \varepsilon_9
\end{aligned}
$$

with ε_i, $i \in \{2, 3, 6, 8, 9\}$ being white noise with mean zero and variance 0.01. Hence, the natural partition of the variables is: $\{\{Z_1, Z_2, Z_3\}, \{Z_4, Z_5, Z_6\}, \{Z_7, Z_8, Z_9\}\}$. Each simulation is run over one hundred repetitions with sample sizes of {50, 100, 500} and the accuracy is assessed in each repetition by computing the fraction of correctly grouped variables. From Table 1, one observes that the dCor based measures are able to correctly group all the variables. This is to be expected as dCor can detect any kind of association between variables, whereas the product moment correlation measures can only detect linear associations. Also, remark that the loss of efficiency caused by the wrapping transformation seems to be minimal in the setting when there are no outliers.

To investigate how the clustering methods perform under cellwise contamination, we follow Van Aelst et al. (2012) and independently add a fraction ε of univariate outliers in d components. We consider $d = 3, 6, 9$, where for each contaminated component Z_l, the outliers are generated from a univariate normal distribution with mean $k\max\{Z_l\}/\sqrt{d}$ and standard deviation 0.1. The results are shown in Table 2 where we observe that the wrapped association measures perform better than their non-robust counterpart. Furthermore, we observe that the distance at which the outliers are placed has little impact on the wrapped measures, due to the choice of the wrapping function. We also observe the need for a large enough sample size, for the correct clusters to be detected when outliers are present.

6.2. *Portfolio optimisation in a comonotonic setting*

In this simulation setting, our aim is to investigate the performance of the different portfolio optimization methods of section 4 in a commonotic setting of normally distributed returns. As shown in Sass and Thös (2021), such a scenario provides a perfect situation to apply the maxCluster and minCluster portfolios because assets in different clusters are independent and have perfect correlation within the clusters. We consider $N = 20$ assets divided into 8 groups as shown in Table 3. Within the 8 groups, we simulate independent, skewed t random variables $\widetilde{\varepsilon}_1, \ldots, \widetilde{\varepsilon}_8$ and define in group b, $Z_{s,t} = \mu_s + \sigma_s \widetilde{\varepsilon}_{b,t}$, $t = 1, \ldots, T$, $s \in I_b$, $b = 1, \ldots, 8$ with I_b the index set of group b, and for each asset, the mean μ_s, volatility, σ_s, skewness λ and kurtosis ν are randomly chosen from Table 4.

We perform 100 repetitions, each consisting of $T = 1000$ observations and employ a moving window approach consisting of an estimation window of $K = 126$ days (corresponding to roughly 6 months) of data. We evaluate the performance of the portfolio optimization methods for monthly rebalancing ($\nu = 21$ days). In each rebalancing period h ($h = 1, \ldots, H$, with H the number of

Table 2
Fraction of correctly grouped variables with $\varepsilon = 0.2$, varying number of observations (n), number of contaminated variables (d), and outlying distance (k).

Method	n	$d = 3$, $k = 6$	$d = 3$, $k = 64$	$d = 6$, $k = 6$	$d = 6$, $k = 64$	$d = 9$, $k = 6$	$d = 9$, $k = 64$
Cor	50	0.46	0.40	0.49	0.40	0.44	0.37
dCor	50	0.95	0.46	0.66	0.40	0.59	0.40
wCor	50	0.63	0.63	0.61	0.62	0.58	0.64
wdCor	50	**1.00**	**1.00**	**0.89**	**0.89**	**0.81**	**0.81**
Cor	100	0.52	0.39	0.52	0.43	0.52	0.42
dCor	100	0.99	0.51	0.89	0.46	0.77	0.41
wCor	100	0.63	0.67	0.68	0.69	0.66	0.66
wdCor	100	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
Cor	500	0.64	0.43	0.57	0.46	0.58	0.41
dCor	500	**1.00**	0.92	**1.00**	0.57	0.99	0.44
wCor	500	0.73	0.73	0.73	0.74	0.75	0.73
wdCor	500	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**

The bold values represent which method has the best results for the corresponding column.

Table 3
Randomly chosen grouping of the assets into the 8 groups.

	Number of assets				
Number of groups	5	4	3	2	1
8	1	1	2	1	3

Table 4
Parameters for the skewed t variables.

μ_s	−0.45	−0.25	−0.2	−0.15	−0.1	−0.05	0.05	0.08	0.1	0.15	0.2	0.21	0.25	0.3	0.45
σ_s	0.5	0.2	0.3	0.25	0.3	0.2	0.25	0.23	0.2	0.4	0.2	0.25	0.3	0.15	0.5
λ_b	−1	−0.86	−0.71	−0.57	−0.43	−0.29	−0.14	0.	0.14	0.29	0.43	0.57	0.71	0.86	1
ν_b	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75

Table 5
Median performance measures for each portfolio over 100 repetitions, without outliers.

	ASR	MD	TO	SSPW	SoR	Vol
EW	0.02	−13.05	**0.00**	**0.05**	0.0	0.01
HERC_cor	0.02	−19.50	0.84	0.21	0.0	0.01
HERC_dcor	0.02	−19.29	1.02	0.25	0.0	0.01
HERC_wcor	0.01	−20.06	0.83	0.19	0.0	0.01
HERC_wdcor	0.02	−19.34	1.02	0.24	0.0	0.01
HRP_cor	0.02	−10.21	0.32	0.08	0.0	**0.00**
HRP_dcor	0.02	−10.82	0.33	0.08	0.0	**0.00**
HRP_wcor	0.02	−10.56	0.31	0.08	0.0	**0.00**
HRP_wdcor	0.02	−11.50	0.33	0.08	0.0	**0.00**
MV	**0.03**	**−8.49**	0.10	0.14	0.0	**0.00**
minCluster_cor	**0.03**	−14.65	0.13	0.13	0.0	0.01
minCluster_dcor	**0.03**	−13.05	0.14	0.13	0.0	0.01
minCluster_wcor	**0.03**	−14.30	0.12	0.13	0.0	0.01
minCluster_wdcor	**0.03**	−13.05	0.13	0.13	0.0	0.01
maxCluster_cor	0.02	−15.02	0.10	0.13	0.0	0.01
maxCluster_dcor	0.02	−14.85	0.10	0.13	0.0	0.01
maxCluster_wcor	0.02	−14.91	0.10	0.13	0.0	0.01
maxCluster_wdcor	**0.03**	−14.91	0.10	0.13	0.0	0.01

The bold values represent which method has the best results for the corresponding column.

rebalancing periods), starting on day $K + 1$, we use the data from the previous K days to estimate the portfolio weights. Using these weights, we compute the portfolio out-of-sample return in rebalancing period $h + 1$. The estimation window is then rolled over by adding the v observations for the next period and removing the v earliest observations. This procedure is then repeated until the latest observation has been used out-of-sample. Applying this rolling window approach to each portfolio, we obtain $T - K$ out-of-sample returns on which we can compute the performance measures presented in Section 5. Table 5 shows the average performance measures over the 100 repetitions. It shows that the minCluster portfolios have the best adjusted Sharpe ratio and the HERC portfolios have the highest Omega ratio. Apart from the EW portfolio, the HRP portfolios are the least concentrated. We also observe that the minVar and maxCluster portfolios have the lowest turnover.

Next, we assess the performance of the portfolios when cellwise outliers are added to the simulated data. For each of the 20 assets, a proportion $\varepsilon = 0.2$ of outliers are added independently, where for each contaminated component Z_l the outliers are generated from a univariate normal distribution with mean $(-1)^{Be} 64 \max\{Z_l\}/\sqrt{d}$ and standard deviation 0.1. Here, Be is a Bernoulli random variable with probability 0.5 to represent the fact that the outlier can either be a positive or a negative return. From Table 6, we observe that the wrapped Cor and dCor based HERC portfolios have the highest adjusted sharpe ratios, showing the efficacy of the wrapping to deal with cellwise outliers and the HERC as a portfolio construction method.

7. Real data example

In this section, the proposed portfolio optimisation methods will be compared to existing methods for a portfolio that contains a combination of traditional (non CC) and CC assets. The traditional assets, shown in Table 7, comprise multiple asset classes, such as equities, commodities, real estate, fiat currencies and fixed income. All traditional asset data has been sourced from Yahoo Finance, using their Python API.

For the CCs, daily data are collected from CoinMarketCap[1] for the 250 Cryptocurrencies (CCs) with the highest trading volume, as per 02/18/2022. The data consist of seven years of daily prices and trading volumes of the CCs as recorded between 01/01/2015 and

[1] www.coinmarketcap.com.

Table 6
Median performance measures for each portfolio over 100 repetitions, with skewed-t cellwise outliers.

	ASR	MD	TO	SSPW	SoR	Vol
EW	0.06	−52.55	**0.00**	**0.05**	0.02	0.05
HERC_cor	0.33	−22.03	1.28	0.28	**0.05**	0.13
HERC_dcor	0.22	−60.65	1.11	0.31	**0.05**	0.15
HERC_wcor	**0.41**	**−12.06**	0.94	0.27	**0.05**	0.11
HERC_wdcor	0.39	−16.16	1.12	0.30	**0.05**	0.12
HRP_cor	0.02	−13.24	0.18	0.20	0.00	**0.01**
HRP_dcor	0.01	−13.36	0.19	0.19	0.00	**0.01**
HRP_wcor	0.02	−13.00	0.17	0.20	0.00	**0.01**
HRP_wdcor	0.02	−12.53	0.17	0.20	0.00	**0.01**
MV	0.02	−16.77	0.08	0.22	0.00	**0.01**
minCluster_cor	0.07	−105.14	1.06	0.12	0.03	0.09
minCluster_dcor	0.12	−98.64	0.94	0.11	0.03	0.09
minCluster_wcor	0.12	−105.33	0.17	0.13	0.03	0.08
minCluster_wdcor	0.12	−91.74	0.24	0.13	0.04	0.09
maxCluster_cor	0.09	−100.25	1.05	0.12	0.03	0.09
maxCluster_dcor	0.07	−95.18	0.98	0.11	0.03	0.09
maxCluster_wcor	0.10	−151.03	0.25	0.13	0.04	0.09
maxCluster_wdcor	0.11	−111.01	0.40	0.13	0.04	0.09

The bold values represent which method has the best results for the corresponding column.

Table 7
List of traditional assets used in the investment universe.

Name	Ticker	Asset class
EURO STOXX 50	^STOXX50E	Equity
S&P100	^SP100	Equity
NIKKEI225	^N225	Equity
FTSE100	^FTSE	Equity
MSCI ACWI COMMODITY PRODUCER	ACWI	Commodities
GOLD	GC = F	Commodities
FTSE EPRA/NAREIT DEV REITS	FFR	Real Estate
EUR/USD	EURUSD = X	Fiat Currency
GBP/USD	GBPUSD = X	Fiat Currency
CNY/USD	CNYUSD = X	Fiat Currency
YEN/USD	JPYUSD = X	Fiat Currency
USA 10Y Treasuries	^TNX	Fixed income

Table 8
Summary statistics of traditional assets and highest trading volume CCs between 2020 and 2022.

	mean	std	min	25%	50%	75%	max	var	skew	kurt	out_frac
^STOXX50E	0.00	0.02	−0.12	−0.01	0.00	0.01	0.09	0.00	−1.05	12.20	0.10
^SP100	0.00	0.02	−0.12	−0.00	0.00	0.01	0.10	0.00	−0.52	12.91	0.10
^N225	0.00	0.01	−0.06	−0.01	0.00	0.01	0.08	0.00	0.25	5.05	0.05
^FTSE	0.00	0.01	−0.11	−0.01	0.00	0.01	0.09	0.00	−0.89	11.51	0.10
ACWI	0.00	0.02	−0.11	−0.00	0.00	0.01	0.08	0.00	−1.14	13.32	0.08
GC = F	0.00	0.01	−0.05	−0.00	0.00	0.01	0.06	0.00	−0.20	4.57	0.08
FFR	0.00	0.02	−0.15	−0.01	0.00	0.01	0.08	0.00	−1.75	17.20	0.09
EURUSD = X	0.00	0.00	−0.03	−0.00	0.00	0.00	0.01	0.00	−0.43	4.18	0.02
GBPUSD = X	0.00	0.01	−0.04	−0.00	0.00	0.00	0.03	0.00	−0.47	6.24	0.06
CNYUSD = X	0.00	0.00	−0.01	−0.00	0.00	0.00	0.01	0.00	0.02	4.39	0.07
JPYUSD = X	−0.00	0.00	−0.03	−0.00	−0.00	0.00	0.02	0.00	−0.22	5.82	0.04
^TNX	0.00	0.06	−0.29	−0.02	0.00	0.02	0.50	0.00	1.60	18.40	0.06
BTC	0.01	0.04	−0.37	−0.02	0.00	0.03	0.19	0.00	−1.05	11.73	0.07
ETH	0.01	0.06	−0.42	−0.02	0.01	0.04	0.19	0.00	−1.13	8.68	0.05
XRP	0.00	0.07	−0.42	−0.02	0.00	0.03	0.38	0.00	−0.23	8.22	0.11
ADA	0.01	0.07	−0.40	−0.03	0.01	0.04	0.32	0.00	0.11	4.82	0.08
LTC	0.00	0.06	−0.36	−0.02	0.00	0.03	0.21	0.00	−1.05	6.80	0.06
BCH	0.00	0.07	−0.43	−0.03	0.00	0.03	0.52	0.00	0.31	14.27	0.07
BNB	0.01	0.07	−0.42	−0.02	0.00	0.03	0.70	0.01	1.70	22.38	0.07
DOT	0.00	0.06	−0.38	−0.02	0.00	0.02	0.33	0.00	0.50	6.92	0.18
DOGE	0.01	0.19	−0.36	−0.02	−0.00	0.02	3.56	0.04	14.43	262.19	0.15
EOS	0.00	0.07	−0.40	−0.02	0.00	0.03	0.55	0.01	0.55	13.20	0.12

01/01/2022. Table 8 shows the descriptive statistics for the daily returns of the traditional assets, along with the top 10 CCs by trading volume between 2020 and 2022. One observes negative skewness, fat tails, and non-normality. Furthermore, the proportion of observations with absolute robust z-scores in excess of 2.5 is computed. It ranges between 0.02 and 0.18 in this time period, showing a small amount of outliers in each of the assets. Looking at the pairwise association measures in Figs. 2 and 3 detects higher association between the CCs over this period and lower associations between CCs and traditional assets. This is not surprising as the CC market seems to be positively correlated due to herding behavior. Furthermore, this shows the potential diversification benefits that can be obtained from adding CCs to an already broadly diversified portfolio.

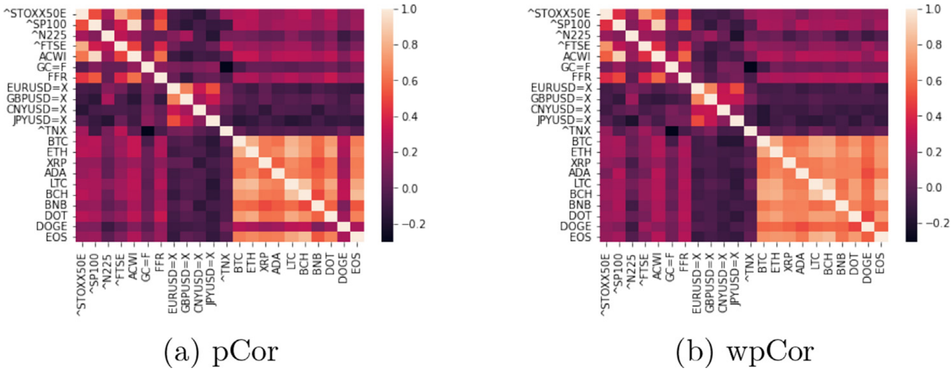

Fig. 2. pCor and wpCor for the top 10 highest CCs with the highest median trading volume between 01/01/2020 and 01/01/2022.

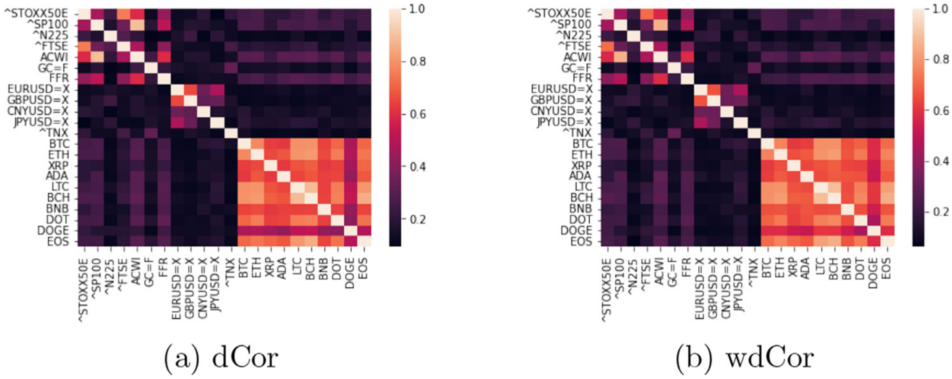

Fig. 3. dCor and wdCor for the top 10 highest CCs with the highest median trading volume between 01/01/2020 and 01/01/2022.

Table 9
Out-of-sample results of portfolio optimization methods on CCs from 01/01/2015 to 01/01/2022.

	ASR	MD	TO	SSPW	OR	Vol
EW	0.08	−26.18	**0.04**	**0.03**	0.02	0.03
MV	0.03	**−11.14**	0.18	0.37	0.00	**0.00**
HRP_cor	0.02	−13.47	0.19	0.30	0.00	**0.00**
HRP_dcor	0.02	−12.52	0.20	0.30	0.00	**0.00**
HRP_wcor	0.03	−11.82	0.18	0.30	0.00	**0.00**
HRP_wdcor	0.03	−11.35	0.19	0.29	0.00	**0.00**
HERC_cor	0.05	−13.05	0.81	0.26	0.01	0.01
HERC_dcor	0.05	−25.84	0.76	0.23	0.01	0.02
HERC_wcor	0.06	−13.58	0.81	0.27	0.00	0.01
HERC_wdcor	0.08	−23.48	0.74	0.22	0.01	0.02
maxCluster_cor	0.06	−28.94	1.14	0.25	**0.03**	0.03
minCluster_cor	0.08	−15.20	0.95	0.26	0.02	0.02
maxCluster_dcor	0.06	−26.99	1.18	0.25	**0.03**	0.03
minCluster_dcor	0.10	−13.61	0.84	0.18	0.02	0.02
maxCluster_wcor	0.05	−32.85	1.17	0.25	0.02	0.03
minCluster_wcor	0.07	−31.17	0.90	0.26	0.02	0.02
maxCluster_wdcor	0.05	−29.62	1.11	0.25	0.02	0.03
minCluster_wdcor	**0.12**	−12.16	0.92	0.17	0.02	0.02

The bold values represent which method has the best results for the corresponding column.

Using the daily returns of the traditional assets together with the highest 250 CCs by trading volume, we apply a similar rolling window approach as in section 6.2 but we also make use of the trading volume information at our disposal. In particular, in each estimation window of $K = 126$ days, we restrict the CCs asset universe to the $M = 20$ most liquid CCs, i.e those with the highest median trading volume. We also impose a no-short selling constraint and impose that the median trading volume should be strictly positive to avoid that the number of constituents of the portfolios be lower than M. The performance measures resulting from the rolling window evaluation with $M = 20$, $\nu = 21$, and $K = 126$ are shown in Table 9. The minCluster portfolios with wrapped distance correlation has the highest adjusted Sharpe ratio and performs 50% better than the EW portfolio, which is a sizeable improvement. Note that the minCluster portfolios also outperforms the maxCluster portfolios.

8. Conclusion

Hierarchical clustering methods have been a recent addition to the universe of portfolio optimization methods (López de Prado, 2016; Raffinot, 2017, 2018; Sass and Thös, 2021). These clustering methods rely on the Pearson correlation to extract similarities between assets and build hierarchies. However, these clustering methods suffer from the presence of cellwise outliers. In this article, we propose the minCluster portfolio which combines cellwise robust association measures, hierarchical clustering and tail risk measures to produce stable portfolio with attractive out-of-sample performance. The attractiveness of the method is assessed via simulation studies and a real data example containing cellwise outliers. The minCluster portfolio outperforms the other portfolio optimization methods in terms of adjusted sharpe ratio. Other portfolio optimizations based on hierarchical clustering also benefit from protection against cellwise outliers.

The proposed association measures are general and could be applied to other (non-hierarchical) clustering methods, to extract clusters in a robust manner. Furthermore, future theoretical research could focus on the robustness properties of the proposed measure. Future applied research could build upon the minCluster portfolio and experiment with different clustering or linkage methods. We note that the proposed methodology for measuring robustness does not consider that outliers in the data may occur in clusters over time. To address this, the current approach uses a moving window for portfolio optimization. However, future research should focus on developing methods that directly account for the possibility of temporal clustering of outliers.

Declaration of competing interest

All authors have none to declare.

References

Alqallaf, F., Aelst, S.V., Yohai, V.J., Zamar, R.H., 2009. Propagation of outliers in multivariate data. Ann. Stat. 37, 311–331.
Artzner, P., Delbaen, F., Eber, J., Heath, D., 1999. Coherent measures of risk. Math. Finance 9, 203–228.
Benhamou, E., Saltiel, D., Ohana, J.J., Atif, J., Laraki, R., 2021. Deep reinforcement learning (drl) for portfolio allocation. In: Dong, Y., Ifrim, G., Mladenić, D., Saunders, C., Van Hoecke, S. (Eds.), Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track. Springer International Publishing, Cham, pp. 527–531.
Boudt, K., Daníelsson, J., Laurent, S., 2013. Robust forecasting of dynamic conditional correlation garch models. Int. J. Forecast. 29, 244–257.
Brauneis, A., Mestel, R., 2019. Cryptocurrency-portfolios in a mean-variance framework. Finance Res. Lett. 28, 259–264.
Burggraf, T., 2021. Beyond risk parity – a machine learning-based hierarchical risk parity approach on cryptocurrencies. Finance Res. Lett. 38, 101523.
Chaouki, A., Hardiman, S., Schmidt, C., Sérié, E., de Lataillade, J., 2020. Deep deterministic portfolio optimization. The Journal of Finance and Data Science 6, 16–30. https://doi.org/10.1016/j.jfds.2020.06.002. https://www.sciencedirect.com/science/article/pii/S2405918820300118.
Chekhlov, A., Uryasev, S., Zabarankin, M., 2005. Drawdown measure in portfolio optimization. Int. J. Theor. Appl. Finance 8, 13–58.
Chen, K., Cheng, T., 2022. Measuring tail risks. The Journal of Finance and Data Science 8, 296–308.
DeMiguel, V., Garlappi, L., Uppal, R., 2009. Optimal versus naive diversification: how inefficient is the 1-n portfolio strategy? Rev. Financ. Stud. 22, 1915–1953.
Elendner, H., Trimborn, S., Ong, B., Lee, T.M., 2018. The cross-section of crypto-currencies as financial assets. In: Lee Kuo Chuen, D., Deng, R. (Eds.), Handbook of Blockchain, Digital Finance, and Inclusion, Volume 1. Academic Press, pp. 145–173.
Gambacciani, M., Paolella, M.S., 2017. Robust normal mixtures for financial portfolio allocation. Econometrics and Statistics 3, 91–111.
Goetzmann, W.N., Kumar, A., 2008. Equity portfolio diversification. Rev. Finance 12, 433–463.
Hafner, C.M., 2018. Testing for bubbles in cryptocurrencies with time-varying volatility. J. Financ. Econom. 18, 233–249.
Hampel, F.R., Rousseeuw, P.J., Ronchetti, E., 1981. The change-of-variance curve and optimal redescending m-estimators. J. Am. Stat. Assoc. 76, 643–648.
Hellmich, M., Kassberger, S., 2011. Efficient and robust portfolio optimization in the multivariate generalized hyperbolic framework. Quant. Finance 11, 1503–1516.
Hoffmann, I., Serneels, S., Filzmoser, P., Croux, C., 2015. Sparse partial robust M regression. Chemometr. Intell. Lab. Syst. 149, 50–59.
Keating, C., Shadwick, W.F., 2002. A universal performance measure. Journal of performance measurement 6, 59–84.
Kojadinovic, I., 2004. Agglomerative hierarchical clustering of continuous variables based on mutual information. Comput. Stat. Data Anal. 46, 269–294.
Liu, W., 2019. Portfolio diversification across cryptocurrencies. Finance Res. Lett. 29, 200–205.
López de Prado, M., 2016. Building diversified portfolios that outperform out of sample. J. Portfolio Manag. 42, 59–69.
Markowitz, H., 1952. Portfolio selection. J. Finance 7, 77–91.
Maronna, R., Martin, D., Yohai, V., 2006. Robust statistics: theory and methods. Wiley Series in Probability and Statistics (Wiley). ISBN: 9780470010921.
Menvouta, E.J., Serneels, S., Verdonck, T., 2022. Sparse dimension reduction based on energy and ball statistics. Advances in Data Analysis and Classification 16, 951–975. https://doi.org/10.1007/s11634-021-00470-7.
Menvouta, E.J., Serneels, S., Verdonck, T., 2023. direpack: a python 3 package for state-of-the-art statistical dimensionality reduction methods. SoftwareX 21, 101282.
Petukhina, A.A., Reule, R.C.G., Härdle, W.K., 2021. Rise of the machines? intraday high-frequency trading patterns of cryptocurrencies. Eur. J. Finance 27, 8–30.
Pézier, J., White, A., 2008. The relative merits of alternative investments in passive portfolios. J. Altern. Investments 10, 37–49.
Pflug, G.C., Pichler, A., Wozabal, D., 2012. The 1/n investment strategy is optimal under high model ambiguity. J. Bank. Finance 36, 410–417.
Pinelis, M., Ruppert, D., 2022. Machine learning portfolio allocation. The Journal of Finance and Data Science 8, 35–54.
Platanakis, E., Sutcliffe, C., Urquhart, A., 2018. Optimal vs naïve diversification in cryptocurrencies. Econ. Lett. 171, 93–96.
Raffinot, T., 2017. Hierarchical clustering-based asset allocation. J. Portfolio Manag. 44, 89–99.
Raffinot, T., 2018. The Hierarchical Equal Risk Contribution Portfolio. https://ssrn.com/abstract=3237540.
Raymaekers, J., Rousseeuw, P.J., 2021. Fast robust correlation for high-dimensional data. Technometrics 63, 184–198.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65.
Sass, J., Thös, A.K., 2021. Risk reduction and portfolio optimization using clustering methods. Econometrics and Statistics (in press).
Serneels, S., Croux, C., Filzmoser, P., Van Espen, P., 2005. Partial robust M-regression. Chemometr. Intell. Lab. Syst. 79, 55–64.
Székely, G., Rizzo, M., 2013. Energy statistics: a class of statistics based on distances. J. Stat. Plann. Inference 143, 1249–1272.
Tibshirani, R., Guenther, W., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. J. Roy. Stat. Soc. Series B 63, 411–423.
Van Aelst, S., Vandervieren, E., Willems, G., 2012. A stahel–donoho estimator based on huberized outlyingness. Comput. Stat. Data Anal. 56, 531–542.
Yu, P., Lee, J.S., Kulyatin, I., Shi, Z., Dasgupta, S., 2019. Model-based Deep Reinforcement Learning for Dynamic Portfolio Optimization. https://doi.org/10.48550/ARXIV.1901.08740.