Finance Research Letters 53 (2023) 103608

Contents lists available at ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

Hierarchical risk parity using security selection based on peripheral assets of correlation-based minimum spanning trees

Younghwan Cho, Jae Wook Song *

Department of Industrial Engineering, Hanyang University, Seoul 04763, Republic of Korea

ARTICLE INFO

Keywords:
Portfolio management
Hierarchical risk parity
Global motion correlation
Minimum spanning tree
Security selection

ABSTRACT

This study proposes hierarchical risk parity portfolios using a new correlation matrix and security selection. We suggest a global motion subtracted correlation matrix, which eliminates the global motion in the cross-correlation matrix. Also, we suggest utilizing the peripheral assets of a correlation-based minimum spanning tree for security selection. The proposed portfolio strategies with security selection outperform benchmarks, showing their nature as smart beta strategies. Specifically, the full correlation with a small number and global motion subtracted correlation with a relatively large number of selected assets exhibit decent performances during the post-crisis bull markets and crisis-induced bear markets, respectively.

1. Introduction

Financial markets dynamically change due to various external and internal factors affecting associated assets. Due to this nature, financial markets have been modeled as complex systems (Mantegna, 1999; Tumminello et al., 2005). Specifically, the correlation-based financial network has been mainly studied via minimum spanning tree (MST) and planar maximally filtered graph (Aste et al., 2010; Di Matteo et al., 2010; Kumar and Deo, 2012; Dai et al., 2016; Song et al., 2019; Ku et al., 2020) in the field of risk management and portfolio optimization (Mastromatteo et al., 2012; Haluszczynski et al., 2017; Wang et al., 2017; Song et al., 2018; Park et al., 2020). In this context, the importance of peripheral nodes in a financial network in constructing a robust portfolio has been discovered (Onnela et al., 2003; Pozzi et al., 2013). However, these studies only considered the mean–variance asset allocation strategy.

Especially, a correlation matrix relies on a reliable statistical estimation. For instance, Pozzi et al. (2013) proposed the shrinkage correlation to compute stable correlation matrices over time, whereas Yan and Zhao (2018) suggested the Gaussian rank correlation to capture the nonlinear relationship between asset returns. Another interesting approach is to extract significant interaction from the matrix. Plerou et al. (1999) claimed that the eigenvalues of the empirical correlation matrix agree with those of the random matrix theory, but few of the largest eigenvalues show significant deviations (Plerou et al., 2002; Utsugi et al., 2004). The largest eigenvalues are considered a global motion that represents the core dynamics of the financial network (Marčenko and Pastur, 1967; Meng et al., 2015; Han et al., 2017; Pan and Sinha, 2007; Borghesi et al., 2007; Jiang et al., 2014). Following the work of Song et al. (2016) and Borghesi et al. (2007), we propose to utilize a new correlation by subtracting the global motion from the full cross-correlation.

From the Markowitz model (Markowitz, 1968) to the risk parity (Qian, 2011), portfolio construction processes lack the hierarchical structure of the financial market. In this regard, De Prado (2016) proposed the new hierarchical risk parity (HRP) model, which outperforms traditional risk minimization portfolios (De Prado, 2016; Burggraf, 2021). However, the HRP requests

* Corresponding author.
E-mail address: jwsong@hanyang.ac.kr (J.W. Song).

https://doi.org/10.1016/j.frl.2022.103608
Received 26 October 2022; Received in revised form 16 December 2022; Accepted 20 December 2022
Available online 23 December 2022
1544-6123/

the complete hierarchy of a financial market, which places a burden on managing a deluge of investments. To cope with this limitation, we propose the HRP model using security selection. Specifically, the proposed portfolio comprises selected assets from the peripheral node pool in correlation-based MST. The correlation matrices utilized for MST are two folds: the full cross-correlation (FC) and global motion subtracted correlation (GMSC). Both correlation matrices maintain or even improve the portfolio performance against benchmarks but different patterns according to the number of selected assets and market conditions.

2. Methods

2.1. Correlation matrix and global motion

Let $P_i(t)$ be the closing price of the ith asset at time t; then the log-return is $R_i(t) = \ln P_i(t) - \ln P_i(t-1)$. Let $\mu(t)$ and $\sigma(t)$ be the arithmetic mean and standard deviation of the log-returns of the moving window length T, $\{R_i(t-T+1), \dots, R_i(t)\}$, respectively; then the normalized return, $\tilde{R}_i(t)$, is

$$\tilde{R}_i(t) = \frac{R_i(t) - \mu(t)}{\sigma(t)}. \tag{1}$$

Since we can define the normalized log-return vector for N number of assets as $\tilde{\mathbf{R}}_i^t = [\tilde{R}_i(t-T+1), \dots, \tilde{R}_i(t)]$ where $i = 1, 2, \dots, N$, the FC at time t can be defined as follows:

$$\rho_{ij}^t = (1/T)(\tilde{\mathbf{R}}_i^t \cdot \tilde{\mathbf{R}}_j^t) - (1/T^2)(\tilde{\mathbf{R}}_i^t \cdot \mathbf{1}_T)(\tilde{\mathbf{R}}_j^t \cdot \mathbf{1}_T) \tag{2}$$

where $\mathbf{1}_T$ indicates a vector of ones with a length of time window T. Note that $\mathbf{x} \cdot \mathbf{y}$ denotes inner product between two arbitrary vectors $\mathbf{x}$ and $\mathbf{y}$, respectively.

Let $\boldsymbol{\rho}$ be the FC matrix of any moving window whose element is ρ_{ij} in Eq. (2). First, the Marchenko–Pastur distribution represents the eigenvalue distribution of a random matrix (Marčenko and Pastur, 1967; Götze and Tikhomirov, 2004). According to the Marchenko–Pastur distribution, the eigenvalues of a random matrix have a theoretical upper bound. Therefore, among the eigenvalues of an empirical correlation matrix, the eigenvectors corresponding to the values exceeding the upper bound can be regarded as not an interaction caused at random. From this point of view, the probability density functions of the eigenvalues of an empirical FC matrix can be divided into three components as follows:

$$\boldsymbol{\rho} = \lambda_1 \mathbf{u}_1 \mathbf{u}_1{}^T + \Sigma_{i=2}^{N_g} \lambda_i \mathbf{u}_i \mathbf{u}_i{}^T + \Sigma_{i=N_g+1}^{N} \lambda_i \mathbf{u}_i \mathbf{u}_i{}^T \tag{3}$$

where λ_i and $\mathbf{u_i}$ denote the list of eigenvalues sorted in descending order of $\boldsymbol{\rho}$ and corresponding eigenvector, respectively. N_g indicates the number of eigenvalues that exceed the upper bound of the theoretical eigenvalue distribution. The theoretical upper bound and N_g are time-varying as the moving window slides. Especially, the first component, $\lambda_1 \mathbf{u_1} \mathbf{u_1}^T$, can be considered a global motion. Note that the global motion refers to the collective reaction of the entire market to outside information (Pan and Sinha, 2007; Laloux et al., 1999; Plerou et al., 2002). Finally, the GMSC matrix, $\tilde{\boldsymbol{\rho}}$, can be defined as follows:

$$\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho} - \lambda_1 \mathbf{u}_1 \mathbf{u}_1{}^T \tag{4}$$

Theoretically, the GMSC matrix focuses on the internal interaction among the idiosyncratic risks of assets by excluding the market risk incurred from the common external information.

2.2. Hierarchical risk parity using peripheral assets in a minimum spanning tree

The correlation matrix can be mapped to a metric space using the distance measure D such that,

$$D_{ij} = \sqrt{2(1 - r_{ij})} \tag{5}$$

where r_{ij} is the elements of either the FC or GMSC matrix. The MST based on the distances in Eq. (5) is a spanning tree whose sum of distances among assets is the smallest. Note that Kruskal's algorithm is employed to compute the MST (Kruskal, 1956).

HRP utilizes hierarchical agglomerative clustering (HAC) to incorporate the market structure in asset allocation. Interestingly, MST and HAC with a single linkage criterion are identical (Mantegna and Stanley, 1999; Marti et al., 2021), whose algorithm is described in Appendix A (Müllner, 2011; Berthold and Höppner, 2016). Thus, HRP asset allocation and MST-driven security selection can be combined into a single portfolio strategy. Since the FC and GMSC used to generate MST are time-varying, the portfolio strategy proposed in this study incorporates the hierarchical structure of a financial market at each rebalancing.

Once tree clustering is performed for HAC, quasi-diagonalization is applied. The covariance matrix of non-normalized log-return is re-arranged to place similar assets closer and dissimilar assets far apart. The single linkage HAC produces a $(N-1) \times 4$ matrix where each row of the linkage matrix may contain clusters or constituents. Starting from the last row of the matrix, we replace the clusters in the row with their constituents recursively until no cluster remains. Finally, the rows and columns of the original covariance matrix are sorted according to the order of single linkage HAC values.

Then we utilize the recursive bisection proposed in De Prado (2016) to compute the portfolio weights as described in Appendix B. Specifically, the recursive bisection starts from the final cluster. Note that the initial weight is set to one for all assets. The cluster encompassing all assets is split into two sub-clusters by moving down the tree structure. The splitting divides the list of investments

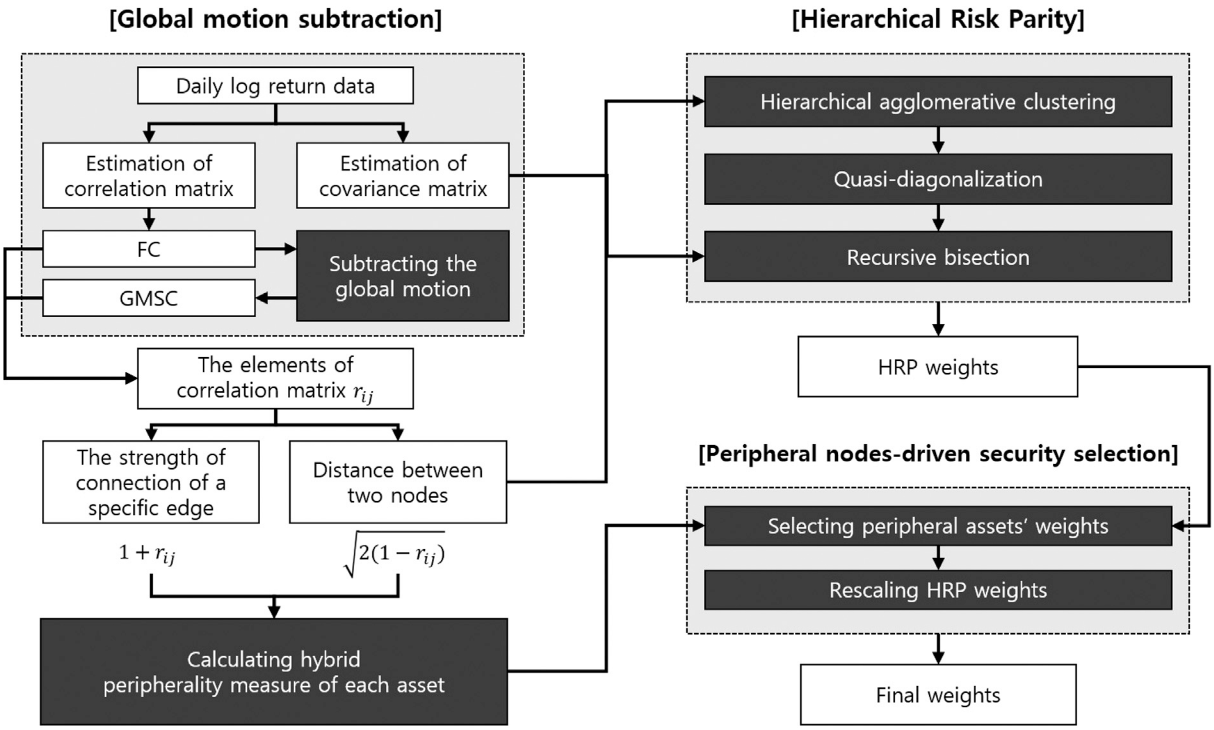


Fig. 1. Overall HRP portfolio building process.

of the upper cluster into two lists proportionally. Then, the variance of each cluster is calculated as the variance of the inverse variance allocation portfolio for each sub-cluster. From each aggregated variance, we compute the split factor and use the factor to re-scale the asset allocation of each subset.

To reduce the number of assets in the portfolio, we select assets using a modified hybrid peripherality measure in Pozzi et al. (2008, 2013), as described in Appendix C. The higher the hybrid measure, the more peripheral the node is. Therefore, we can select the κ number of outermost peripheral assets with the highest hybrid measure values.

The last step of portfolio management is re-scaling HRP weights for selected assets. Our objective is to produce the final HRP weights that maintain the characteristics of the entire hierarchy of the financial market. Thus, we divide the weights of each selected asset by the sum of all selected assets' weights. Finally, the overall process of building HRP portfolios for FC and GMSC matrices using security selection can be summarized as shown in Fig. 1.

3. Empirical results

3.1. Data & backtesting set-up

We utilize the daily closing prices of S&P500 constituents from July 2000 to August 2021 (5320 observations). We use the rolling window of 120 days to estimate the correlation matrix. Thus, we only consider assets with at least 120 log-returns at rebalancing. The portfolio is rebalanced monthly on the first trading day. Thus, the first day of backtesting is January 2, 2001, and rolled forward until August 31, 2021. The number of assets in S&P500 began with 497 in January 2001 and ended with 504 in August 2021. We evaluate the performance of the proposed portfolio strategy for $\kappa = \{5, 10, 20, 30, 50, 100, 200, 300\}$ against those of the S&P500 index and HRP strategy using all assets.

3.2. Correlation matrix & minimum spanning tree

Fig. 2 shows the probability density function of the eigenvalues of an empirical FC matrix on August 1, 2021. Some eigenvalues of the FC matrix are significantly outside the upper limit of the Marchenko–Pastur distribution. Furthermore, the blue bar represents the eigenvalue distribution of a surrogate correlation matrix estimated by the time-shuffled log-return series of each asset Pan and Sinha (2007). We confirm that none of the eigenvalues of the surrogate correlation matrix exceeds the theoretical upper limit of the eigenvalues of the random matrix.

FC, GMC, and GMSC matrices on August 1, 2021, are shown in Fig. 3. The FC matrix with quasi-diagonalization in Fig. 3(a) shows many significant clusters as blue squares along the diagonal line. The GMC in Fig. 3(b) reveals the global motion with market leading

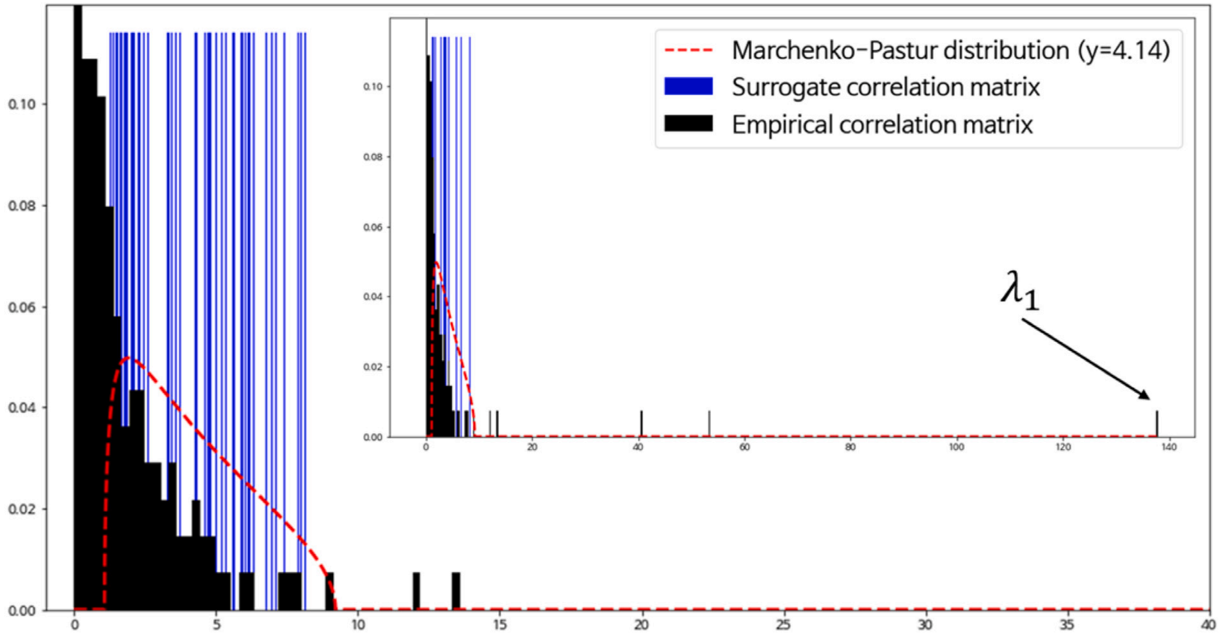
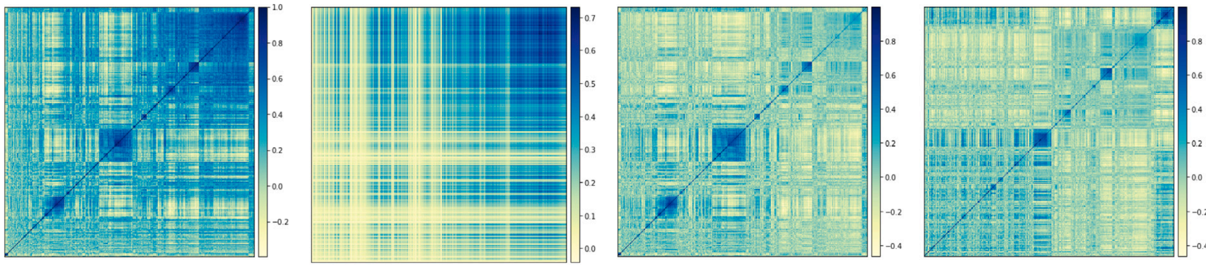


Fig. 2. The probability density function of the eigenvalues from S&P500 on August 1, 2021.



(a) FC with quasi-diagonalization **(b)** GMC from FC **(c)** GMSC **(d)** GMSC with quasi-diagonalization

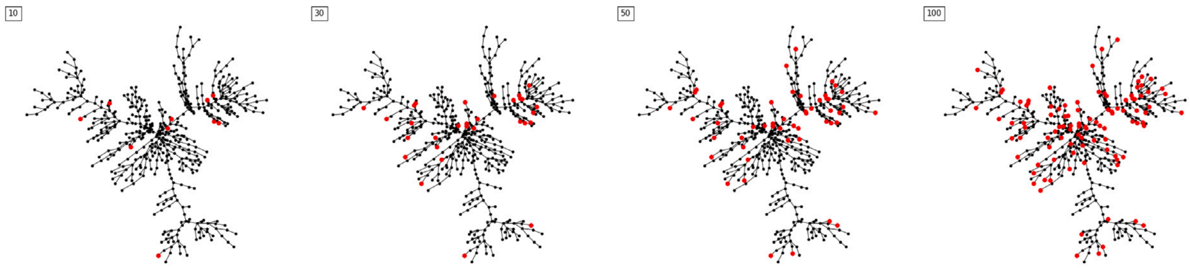
Fig. 3. Snapshots of full cross-correlation, global motion, global motion subtracted correlation matrices on August 1, 2021. Correlation coefficients are described with the color scheme. The higher the correlation, the darker the blue the lattice appears.

assets in the upper right corner. Fig. 3(c) shows fewer clusters and lower correlations. Therefore, the GMSC stresses the importance of assets whose return is less sensitive to market conditions. Then, the quasi-diagonalization is applied to GMSC as in Fig. 3(d).

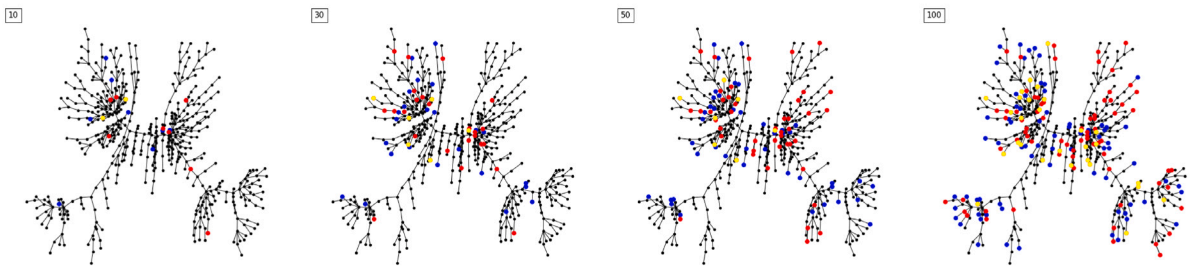
We further analyze the structure of the financial market and selected peripheral assets using the sample configurations of MST in Fig. 4. Figs. 4(a)–4(d) and 4(e)–4(h) show the MSTs of FC and GMSC on August 1, 2021, for 10, 30, 50, and 100 selected peripheral assets, respectively. The MST of the FC shows a radial hierarchical structure with the most connected asset at the center. The MST of GMSC is relatively decentralized by removing the most central asset whose price dynamics is expected to lead the financial market, providing sub-divided hierarchical structure. In addition, all peripheral assets selected in FC (red) and GMSC (blue) for different κ are apart from the center of MST with few connections. The small number of simultaneously selected assets (yellow) implies the difference in security selection between FC and GMSC.

3.3. Portfolio performance

For the entire period, HRP portfolio performances based on FC (FC-HRP) and GMSC (GMSC-HRP) matrices for $\kappa = \{5, 10, 20, 30, 50, 100, 200, 300\}$ are evaluated against benchmarks. Note that FC-HRP with all assets is the HRP model proposed in De Prado (2016). The backtesting considers the 10bps transaction cost and removed assets from the S&P500. At first, the cumulative portfolio log-returns in Fig. 5 shows that FC- and GMSC-HRP outperform S&P500 except for GMSC-HRP with a small κ. Furthermore, FC- and GMSC-HRP with all assets are not the highest cumulative log-return portfolios, implying the effectiveness of the proposed security selection. The trends in cumulative log-returns from FC- and GMSC-HRP are almost identical to that of S&P500 in general. Note that the similarity in trend weakens as κ decreases. Even with a similar trend, FC- and GMSC-HRP show higher portfolio performance than



(a) FC ($\kappa = 10$) **(b)** FC ($\kappa = 30$) **(c)** FC ($\kappa = 50$) **(d)** FC ($\kappa = 100$)



(e) GMSC ($\kappa = 10$) **(f)** GMSC ($\kappa = 30$) **(g)** GMSC ($\kappa = 50$) **(h)** GMSC ($\kappa = 100$)

Fig. 4. Snapshots of the FC and GMSC minimum spanning trees for $\kappa = 10, 20, 30, 50, 100$ on August 1, 2021. The red and blue nodes indicate the selected peripheral assets from the FC and GMSC, respectively. The yellow nodes in the GMSC minimum spanning tree indicate the selected peripheral assets from the FC and GMSC simultaneously.

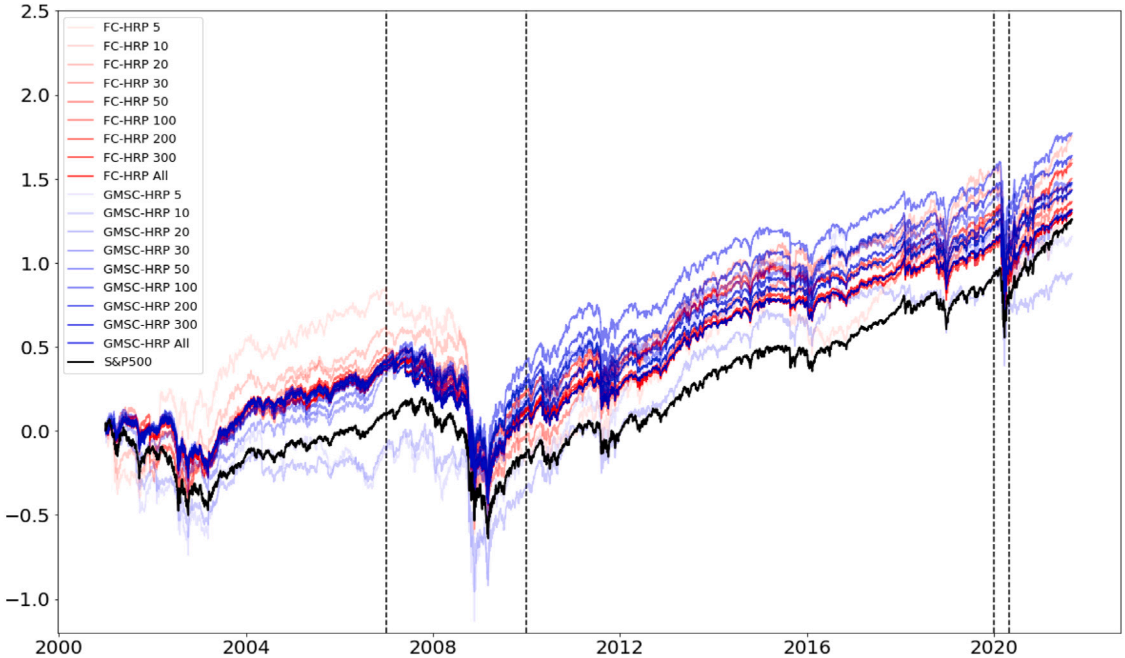


Fig. 5. Cumulative log-returns of HRP models for the total backtesting period. The red and blue solid lines indicate the FC- and GMSC-HRP models, respectively. The black vertical dotted lines indicate the delimiter for different sub-periods.

the benchmarks. It implies that FC- and GMSC-HRP with security selection are smart beta strategies that seek to improve returns and reduce risk using the periphery of assets in the financial market structure as a factor.

Table 1
Performance of each portfolio for the total backtesting period.

		$\kappa = 5$	$\kappa = 10$	$\kappa = 20$	$\kappa = 30$	$\kappa = 50$	$\kappa = 100$	$\kappa = 200$	$\kappa = 300$	All	S&P500
C/R	FC-HRP	0.1282[a]	0.2319[a]	0.1956[a]	0.1560	0.1693	0.1900	0.1411	0.1329	0.1290	0.1224
	GMSC-HRP	0.1056	0.0751	0.1318	0.1618[a]	0.2368[a]	0.2008[a]	0.1638[a]	0.1546[a]	0.1317[a]	
S/D	FC-HRP	0.2868	0.2435	0.2235	0.2148[a]	0.2030[a]	0.1921[a]	0.1856[a]	0.1790[a]	0.1716	0.1961
	GMSC-HRP	0.2752[a]	0.2434[a]	0.2231[a]	0.2194	0.2117	0.2036	0.1932	0.1859	0.1705[a]	
D/D	FC-HRP	0.2112	0.1794	0.1685	0.1608[a]	0.1510	0.1412[a]	0.1364[a]	0.1315[a]	0.1260	0.1423
	GMSC-HRP	0.1992[a]	0.1759[a]	0.1614[a]	0.1589	0.1534[a]	0.1475	0.1404	0.1352	0.1251[a]	
MDD	FC-HRP	−0.8105	−0.7058[a]	−0.6589	−0.6726[a]	−0.6172	−0.5976[a]	−0.6156	−0.6074[a]	−0.6197	−0.6103
	GMSC-HRP	−0.7779[a]	−0.7173	−0.6393[a]	−0.6865	−0.6018[a]	−0.6239	−0.6048[a]	−0.6080	−0.6127[a]	
SH/R	FC-HRP	0.4471[a]	0.9522[a]	0.8752[a]	0.7260	0.8340	0.9891[a]	0.7602	0.7422	0.7517	0.6241
	GMSC-HRP	0.3838	0.3083	0.5907	0.7374[a]	1.1182[a]	0.9863	0.8479[a]	0.8312[a]	0.7723[a]	
SO/R	FC-HRP	0.6073[a]	1.2929[a]	1.1607[a]	0.9697	1.1213	1.3451	1.0338	1.0102	1.0242	0.8597
	GMSC-HRP	0.5301	0.4267	0.8164	1.0184[a]	1.5438[a]	1.3615[a]	1.1667[a]	1.1429[a]	1.0528[a]	
CA/R	FC-HRP	0.1582[a]	0.3286[a]	0.2969[a]	0.2319	0.2743	0.3179	0.2291	0.2188	0.2082	0.2005
	GMSC-HRP	0.1358	0.1046	0.2061	0.2357[a]	0.3934[a]	0.3218[a]	0.2709[a]	0.2542[a]	0.2150[a]	
P/Y	FC-HRP	14[a]	16[a]	16[a]	15	14	16	15	16	16	16
	GMSC-HRP	13	13	14	15	16[a]	16	15	16	16	
M/TO	FC-HRP	1.8092[a]	1.7497[a]	1.7043[a]	1.6375[a]	1.5294[a]	1.2997[a]	0.9484[a]	0.7155	0.3091	
	GMSC-HRP	1.9233	1.8684	1.8032	1.7548	1.6346	1.3762	0.9763	0.7084[a]	0.2987[a]	
SD/TO	FC-HRP	0.2962	0.2729	0.2412	0.2333	0.2306	0.2163	0.1901	0.1541	0.0908[a]	
	GMSC-HRP	0.1726[a]	0.1783[a]	0.1603[a]	0.1550[a]	0.1979[a]	0.1974[a]	0.1769[a]	0.1420[a]	0.0930	

Note: The abbreviations C/R, S/D, D/D, MDD, SH/R, SO/R, CA/R, P/Y, M/TO, and SD/TO refer to the (annualized) compound return, standard deviation, downside deviation, maximum drawdown, Sharpe ratio, Sortino ratio, Calmar ratio, profitable years, mean of monthly turnover, and standard deviation of monthly turnover, respectively.

[a]Indicates the higher portfolio performance between the FC- and GMSC-HRP.

The detailed portfolio performances are summarized in Table 1. When all assets are selected, FC- and GMSC-HRP are superior to S&P500 in most of performance measures. FC-HRP using security selection for all κ except for five assets obtains a higher annualized compound return (C/R) than all assets. Similarly, GMSC-HRP using security selection for all κ except for five and ten assets generates higher C/R than all assets and S&P500. FC- and GMSC-HRP show less volatility with smaller standard deviations (S/D) and downside deviations (D/D) than that of S&P500 even though the number of assets is reduced to 100 and 200, respectively. FC-HRP with $50 \leq \kappa \leq 300$ and GMSC-HRP with $\kappa = 50, 200, 300$ show smaller maximum drawdowns (MDD) than all assets.

Interestingly, FC- and GMSC-HRP with security selection show higher returns and comparable volatility against all assets and S&P500. Such inference is supported by the risk-adjusted returns: Sharpe Ratio (SH/R), Sortino Ratio (SO/R), and Calmar Ratio (CA/R). From the SH/R and SO/R perspective, the best strategy is GMSC-HRP with $\kappa = 50$. Most FC- and GMSC-HRP with security selection show higher SH/R and SO/R than all assets and S&P500. For CA/R, FC-HRP is superior to benchmarks for all κ except $\kappa = 5$, while GMSC-HRP is superior to all assets and S&P500 when $30 \leq \kappa \leq 300$ and $20 \leq \kappa \leq 300$, respectively. Therefore, FC- and GMSC-HRP are superior to S&P500, and the security selection based on peripheral assets of MST can improve the performance of HRP models.

Furthermore, we compare FC- and GMSC-HRP in terms of security selection. At first, FC-HRP with $\kappa = 5, 10, 20$ outperforms GMSC-HRP in C/R, risk-adjusted returns, and profitable years (P/Y). However, GMSC-HRP shows a lower portfolio risk in S/D, D/D, and MDD. When $30 \leq \kappa \leq 300$, GMSC-HRP outperforms FC-HRP in C/R and risk-adjusted returns, but P/Y and MDD are comparable for both portfolios. The monthly turnover mean (M/TO) and the standard deviation (SD/TO) are calculated as described in Gu et al. (2020). M/TO is two if all assets in the portfolio are replaced in rebalancing, whereas the value is zero when no asset is replaced. Note that higher turnover yields higher transaction costs. M/TO is smaller in FC-HRP, but SD/TO is smaller in GMSC-HRP. In conclusion, FC-HRP is suitable for individual investors whose portfolio has limited capital with a modest number of assets, whereas GMSC-HRP is suitable for institutional investors who consider large capital and abundant assets for portfolio management.

Portfolio performances are further investigated for different sub-periods, which include several bull, sideways, and bear markets. The total period is divided into five sub-periods: sideways market from 01/2001 to 12/2006, sub-prime mortgage crisis from 01/2007 to 12/2009, a long-term bull market after the crisis from 01/2010 to 12/2019, COVID-19 pandemic crisis from 01/2020 to 04/2020, and a bull market after the pandemic from 05/2020 to 08/2021. Portfolio performances are summarized in Table 2.

In the first sub-period, a sideways market, both portfolios with all assets outperform the S&P500 in all performance measures. The performances of both portfolios are higher than that of S&P500 even when the number of assets is reduced to 50. It implies that the proposed security selection based on peripheral assets of MST is effective in the sideways market. In particular, FC-HRP with $5 \leq \kappa \leq 100$ and GMSC-HRP with $\kappa = 50, 100, 200$ record higher C/R than each HRP with all assets. Both portfolios show less portfolio risk than S&P500 when $50 \leq \kappa \leq 300$. FC-HRP shows a higher SH/R than all assets and S&P500 for most κ, whereas GMSC-HRP fails to outperform all assets. Specifically, FC-HRP is superior to GMSC-HRP when $\kappa \leq 30$, which follows the performance in the total period.

In the second sub-period, the Subprime mortgage crisis, FC- and GMSC-HRP with all assets fail to outperform the S&P500. However, both portfolios with security selection outperform the S&P500 in terms of portfolio risk. For C/R, GMSC-HRP outperforms the S&500 for all κ, whereas FC-HRP only outperforms with $\kappa = 100$. In essence, the proposed security selection improves the portfolio performance during the Sub-prime mortgage crisis where the proposed GMSC-based MST shows higher performance than FC.

In the third sub-period, a bull market after the financial crisis, both portfolios of all assets outperform the S&P500 in all performance measures except the C/R, whereas both portfolios with security selection even outperform the C/R of S&P500.

Table 2
Sub-period portfolio performance analysis.

		$\kappa = 5$	$\kappa = 10$	$\kappa = 20$	$\kappa = 30$	$\kappa = 50$	$\kappa = 100$	$\kappa = 200$	$\kappa = 300$	All	S&P500
Sub-period 1: 01/2001–12/2006 (Sideways market)											
C/R	FC-HRP	0.2272[a]	0.1430[a]	0.1086[a]	0.0887[a]	0.0864	0.0946[a]	0.0747	0.0727	0.0822[a]	0.0176
	GMSC-HRP	−0.0222	−0.0167	0.0655	0.0752	0.0870[a]	0.0849	0.0838[a]	0.0740[a]	0.0784	
S/D	FC-HRP	0.2320	0.1905	0.1956	0.1815	0.1635	0.1492[a]	0.1445[a]	0.1394[a]	0.1354[a]	0.1707
	GMSC-HRP	0.2307[a]	0.1859[a]	0.1678[a]	0.1608[a]	0.1569[a]	0.1536	0.1472	0.1448	0.1356	
MDD	FC-HRP	−0.3627[a]	−0.3249[a]	−0.4818	−0.4615	−0.3920	−0.3464	−0.3107	−0.3126[a]	−0.2943[a]	−0.4605
	GMSC-HRP	−0.5785	−0.5176	−0.4087[a]	−0.3634[a]	−0.3490[a]	−0.3380[a]	−0.3081[a]	−0.3163	−0.3000	
SH/R	FC-HRP	0.9794[a]	0.7504[a]	0.5550[a]	0.4888[a]	0.5284	0.6339[a]	0.5173	0.5217[a]	0.6073[a]	0.1031
	GMSC-HRP	−0.0961	−0.0899	0.3902	0.4679	0.5544[a]	0.5528	0.5693[a]	0.5111	0.5781	
CA/R	FC-HRP	0.6263[a]	0.4400[a]	0.2254[a]	0.1922[a]	0.2204	0.2730[a]	0.2406	0.2326	0.2793[a]	0.0382
	GMSC-HRP	−0.0383	−0.0323	0.1602	0.2070	0.2492[a]	0.2512	0.272[a]	0.234[a]	0.2613	
Sub-period 2: 01/2007–12/2009 (Sub-prime mortgage crisis)											
C/R	FC-HRP	−0.2082	−0.1414	−0.1059	−0.1235	−0.0847	−0.0651	−0.0723	−0.0723	−0.0909	−0.0710
	GMSC-HRP	−0.0019[a]	−0.0686[a]	−0.0425[a]	−0.0668[a]	−0.0048[a]	−0.0323[a]	−0.0406[a]	−0.0609[a]	−0.0867[a]	
S/D	FC-HRP	0.4384[a]	0.3719[a]	0.3243[a]	0.3200[a]	0.3024[a]	0.2917[a]	0.2823[a]	0.2744[a]	0.2672	0.2995
	GMSC-HRP	0.4503	0.4001	0.3481	0.3515	0.3343	0.3179	0.2994	0.2873	0.2651[a]	
MDD	FC-HRP	−0.7923	−0.7056	−0.6552	−0.6726[a]	−0.6172	−0.5976[a]	−0.6156	−0.6074[a]	−0.6197	−0.6103
	GMSC-HRP	−0.7216[a]	−0.6810[a]	−0.6393[a]	−0.6865	−0.6018[a]	−0.6239	−0.6048[a]	−0.6080	−0.6127[a]	
SH/R	FC-HRP	−0.4749	−0.3802	−0.3265	−0.3861	−0.2801	−0.2234	−0.2560	−0.2635	−0.3403	−0.2372
	GMSC-HRP	−0.0043	−0.1714	−0.1221	−0.1900	−0.0143	−0.1017	−0.1357	−0.2121	−0.3271	
CA/R	FC-HRP	−0.2628	−0.2004	−0.1616	−0.1837	−0.1372	−0.1090	−0.1174	−0.119	−0.1467	−0.1164
	GMSC-HRP	−0.0027	−0.1007	−0.0665	−0.0973	−0.008	−0.0518	−0.0672	−0.1002	−0.1416	
Sub-period 3: 01/2010–12/2019 (Bull market)											
C/R	FC-HRP	0.2841[a]	0.3383[a]	0.2672[a]	0.2530[a]	0.2165[a]	0.1915	0.1614	0.1736	0.1788	0.1855
	GMSC-HRP	0.2592	0.2507	0.1733	0.2414	0.2138	0.2096[a]	0.1685[a]	0.1843[a]	0.1814[a]	
S/D	FC-HRP	0.2204	0.1794[a]	0.1620[a]	0.1593[a]	0.1511[a]	0.1438[a]	0.1397[a]	0.1330[a]	0.1259	0.1479
	GMSC-HRP	0.2018[a]	0.1834	0.1722	0.1701	0.1613	0.1571	0.1473	0.1410	0.1257[a]	
MDD	FC-HRP	−0.3307	−0.2557[a]	−0.2564[a]	−0.2441[a]	−0.2630	−0.2394[a]	−0.2171[a]	−0.2081[a]	−0.1849	−0.2070
	GMSC-HRP	−0.3180[a]	−0.2809	−0.2710	−0.2732	−0.2509[a]	−0.2495	−0.2463	−0.2271	−0.1831[a]	
SH/R	FC-HRP	1.2892[a]	1.8851[a]	1.6500[a]	1.5888[a]	1.4330[a]	1.3324	1.1555[a]	1.3055	1.4208	1.2546
	GMSC-HRP	1.2847	1.3666	1.0061	1.4194	1.3257	1.3341[a]	1.1443	1.3074[a]	1.4431[a]	
CA/R	FC-HRP	0.8592[a]	1.3229[a]	1.0421[a]	1.0367[a]	0.8231	0.8001	0.7435[a]	0.8343[a]	0.9670	0.8962
	GMSC-HRP	0.8151	0.8924	0.6394	0.8835	0.8521[a]	0.8402[a]	0.6842	0.8118	0.9907[a]	
Sub-period 4: 01/2020–04/2020 (COVID-19 pandemic crisis)											
C/R	FC-HRP	−0.8415	−0.8528	−0.7915	−0.6895	−0.6395	−0.5317	−0.4609	−0.4609	−0.4169	−0.3258
	GMSC-HRP	−0.5311[a]	−0.5091[a]	−0.5988[a]	−0.5379[a]	−0.5860[a]	−0.4944[a]	−0.4126[a]	−0.3715[a]	−0.3795[a]	
S/D	FC-HRP	0.7920	0.7520	0.7232	0.6656	0.6584	0.6249	0.5914[a]	0.5789	0.5424	0.5391
	GMSC-HRP	0.7295[a]	0.6822[a]	0.6830[a]	0.6482[a]	0.6545[a]	0.6217[a]	0.6021	0.5699[a]	0.5310[a]	
MDD	FC-HRP	−0.5675	−0.5432	−0.5109	−0.4842	−0.4754[a]	−0.4515	−0.4197	−0.4150	−0.3927	−0.3610
	GMSC-HRP	−0.4759[a]	−0.4716[a]	−0.4781[a]	−0.4595[a]	−0.4770	−0.4382[a]	−0.4161[a]	−0.3963[a]	−0.3801[a]	
SH/R	FC-HRP	−1.0624	−1.1340	−1.0943	−1.0358	−0.9714	−0.8509	−0.7793	−0.7962	−0.7686	−0.6044
	GMSC-HRP	−0.7280	−0.7462	−0.8767	−0.8299	−0.8952	−0.7952	−0.6852	−0.6518	−0.7147	
CA/R	FC-HRP	−1.4828	−1.5698	−1.5493	−1.4239	−1.3452	−1.1778	−1.0981	−1.1107	−1.0616	−0.9025
	GMSC-HRP	−1.1160	−1.0795	−1.2523	−1.1708	−1.2283	−1.1283	−0.9914	−0.9372	−0.9983	
Sub-period 5: 05/2020–08/2021 (Bull market)											
C/R	FC-HRP	0.3915[a]	0.6137[a]	0.5231[a]	0.4802[a]	0.4746[a]	0.4961[a]	0.4421[a]	0.3486[a]	0.3196[a]	0.4483
	GMSC-HRP	0.1893	0.1702	0.3137	0.2685	0.4307	0.3660	0.3433	0.3377	0.3095	
S/D	FC-HRP	0.2765	0.2442	0.1932	0.1880	0.1830	0.1678	0.1643	0.1548[a]	0.1460	0.1623
	GMSC-HRP	0.2043[a]	0.1893[a]	0.1859[a]	0.1753[a]	0.1766[a]	0.1667[a]	0.1640[a]	0.1616	0.1456[a]	
MDD	FC-HRP	−0.1472	−0.1244[a]	−0.0963[a]	−0.0902[a]	−0.0902[a]	−0.0928[a]	−0.1013[a]	−0.1002[a]	−0.0965[a]	−0.0979
	GMSC-HRP	−0.1244[a]	−0.1333	−0.1338	−0.1220	−0.1269	−0.1168	−0.1159	−0.1121	−0.0967	
SH/R	FC-HRP	1.4156[a]	2.513[a]	2.708[a]	2.5548[a]	2.5938[a]	2.9575[a]	2.6901[a]	2.2519[a]	2.1895[a]	2.7626
	GMSC-HRP	0.9269	0.8995	1.6873	1.5316	2.4386	2.1962	2.0924	2.090	2.1266	
CA/R	FC-HRP	2.6598[a]	4.9332[a]	5.4315[a]	5.3226[a]	5.2609[a]	5.346[a]	4.3624[a]	3.4796[a]	3.311[a]	4.5785
	GMSC-HRP	1.5216	1.2767	2.3435	2.2002	3.3935	3.133	2.9616	3.0137	3.2012	

Note: The abbreviations C/R, S/D, MDD, SH/R, and CA/R refer to the (annualized) compound return, standard deviation, maximum drawdown, Sharpe ratio, and Calmar ratio, respectively.

[a]Indicates the higher portfolio performance between the FC- and GMSC-HRP.

Furthermore, with relatively low portfolio risk, FC- and GMSC-HRP with most of κ outperform the S&P500 in SH/R. Compared to all assets, both portfolios with small κ show improvement in SH/R and CA/R. Specifically, a substantially higher SH/R is observed for FC-HRP with $\kappa = 20$ than the two benchmarks. In summary, FC-HRP-based security selection is more effective than GMSC-HRP in a bull market, and better performance can be expected at lower κ.

In the fourth sub-period, a bear market during the COVID-19 pandemic crisis, both portfolios with all assets and security selection fail to outperform the S&P500 for all performance measures. In C/R and MDD, GMSC-HRP is superior to FC-HRP for all κ. Also, the S/D of GMSC is smaller than FC-HRP in all κ except for $\kappa = 200$. Along with the second sub-period, GMSC is more effective in a bear market than FC.

In the last sub-period, a bull market after the pandemic crisis, both portfolios with all assets show lower S/D and MDD than S&P500. FC-HRP with $\kappa = 10, 20, 30, 50, 100$ shows higher C/R than S&P500 and all assets simultaneously. GMSC-HRP shows lower C/R than S&P500, but the security selection for $\kappa = 20, 50, 100, 200, 300$ improves the C/R. Furthermore, FC-HRP substantially outperforms two benchmarks in SH/R and CA/R in smaller kappa. In summary, both portfolios achieve better performance than all assets through security selection. Moreover, as in the result of the third sub-period, FC improves portfolio performance compared to GMSC in the bull market.

4. Conclusion

We propose HRP models using the minimum spanning tree-driven security selection based on two different correlation matrices, FC and GMSC. The proposed portfolios outperform the benchmarks even when the number of selected assets decreases, implying the efficacy of peripheral asset selection. Furthermore, we discover that FC-HRP with few selected assets is superior in recouping portfolio loss during the post-crisis bull market, whereas GMSC-HRP with many selected assets can minimize the portfolio loss during the financial crisis. Therefore, we expect that the proposed HRP model will be of great utility to individual and institutional investors in the practical management of portfolios.

CRediT authorship contribution statement

Younghwan Cho: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Visualization, Validation, Resources, Data curation. **Jae Wook Song:** Conceptualization, Methodology, Investigation, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Data availability

Data will be made available on request.

Acknowledgments

This work has been supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1F1A1062917).

Appendix A. Hierarchical agglomerative clustering with single linkage

Algorithm 1: Pseudocode for hierarchical agglomerative clustering with single linkage

Input :
node labels S
pairwise Euclidean dissimilarities between N assets d
Output :
stepwise dendrogram L which is a (N-1)×4 matrix
Initialization :
Number of input assets N $\leftarrow |S|$; *Output list of N-1 quadruplets L*$\leftarrow$ []; *size[x]* $\leftarrow 1$, $\forall x \in S$
for $i \leftarrow 0, \cdots, N-2$ **do**
    $(a, b) \leftarrow argmin_{S\times S\setminus \triangle} d$
    $S \leftarrow S \setminus \{a, b\}$
    Create a new node label $n \notin S$.
    $size[n] \leftarrow size[a] + size[b]$
    Append (a, b, d[a,b], $size[n]$*) to L.*
    Update d with the distance information d[n,x]=d[x,n] = min(d[a,x], d[b,x]), $\forall x \in S$.
    $S \leftarrow S \cup \{n\}$
end
return L
{$\triangle$ denotes the diagonal in the Cartesian product linked to $S \times S$.}

Appendix B. Bisection of hierarchical risk parity

See Algorithm 2.

Appendix C. Hybrid peripherality measure

Basically, HRP process is performed with the distance measure from Eq. (5). However, the weight between nodes i and j changes to $1+r_{ij}$ when the weighted degree centrality and the weighted eigenvector centrality are computed as described in Fig. 1. Note that the weight is the same as the HRP process in Eq. (5) for the weighted Eccentricity, betweenness centrality, and closeness centrality. The tied ranks of all five measures are calculated to assign higher ranks (close to 1) to central vertices except for closeness centrality. Let R denotes the rank of an asset for a certain measure. The subscripts D, BC, EC, E, and C refer to the degree centrality, betweenness centrality, eigenvector centrality, eccentricity, and closeness centrality, respectively. The superscripts w and uw indicate the weighted and unweighted MST, respectively. Based on the work of Pozzi et al. (2013), a hybrid peripherality measure P can be defined as follows:

$$X = ({R_D}^{w} + {R_D}^{uw} + {R_{BC}}^{w} + {R_{BC}}^{uw} - 4)/4 \times (N-1) \tag{C.1}$$

$$Y = ({R_E}^{w} + {R_E}^{uw} + {R_C}^{w} + {R_C}^{uw} + {R_{EC}}^{w} + {R_{EC}}^{uw} - 6)/6 \times (N-1) \tag{C.2}$$

$$P = X + Y \tag{C.3}$$

where N is the number of assets. Note that the value of P is small for central assets and large for peripheral assets.

Algorithm 2: Pseudocode for recursive bisection of hierarchical risk parity

1 **Input :**
2 *covariance matrix of all assets* V
3 *sorted asset list* l, *with* $l = \{l_0\}$, *with* $l_0 = n_{n=1,\cdots,N}$ *where each integers match with assets' tickers*
4 **Output :**
5 *return weight* $0 \leq w_i \leq 1$, $\forall i = 1 \cdots, N$, *satisfying* $\sum_{i=1}^{N} w_i = 1$
6 **Initialization :**
7 *initial weight* $w_i = 1$, $\forall i = 1 \cdots, N$
8 **if** $|l_i| = 1, \forall l_i \in l$ **then** go to line 16
9 **for** $\forall l_i \in l$ *such that* $|l_i| > 1$ **do**
10 *Bisect* l_i *into two subsets,* $l_i^{(1)} \cup l_i^{(2)} = l_i$, *where* $|l_i^{(1)}| = int[\frac{1}{2}|l_i|]$, *while preserving the order*
11 *Define the variance of each subset* $l_i^{(j)}, j = 1, 2$, *as the quadratic form* $\tilde{V}_i^{(j)} = \tilde{w}_i^{(j)T} V_i^{(j)} \tilde{w}_i^{(j)}$, *where* V_i^j *is the covariance matrix of assets in the* $l_i^{(j)}$ *bisection, and* $\tilde{w}_i^{(j)}$ *is inverse-variance weightings based on* V_i^j
12 *Compute the split factor:* $\alpha_i = 1 - \frac{\tilde{V}_i^{(1)}}{\tilde{V}_i^{(1)} + \tilde{V}_i^{(2)}}$
13 *Re-scale allocations* w_n *by a factor of* $\alpha_i, \forall n \in l_i^{(1)}$, *and*
14 *Re-scale allocations* w_n *by a factor of* $(1 - \alpha_i), \forall n \in l_i^{(2)}$
15 *Loop to line* 8
16 **end**
17 **return** $w_i, \forall i = 1 \cdots, N$
18 $\{T$ denotes the transpose operator of a vector.$\}$

References

Aste, T., Shaw, W., Di Matteo, T., 2010. Correlation structure and dynamics in volatile markets. New J. Phys. 12 (8), 085009.

Berthold, M.R., Höppner, F., 2016. On clustering time series using euclidean distance and pearson correlation. arXiv preprint arXiv:1601.02213.

Borghesi, C., Marsili, M., Micciche, S., 2007. Emergence of time-horizon invariant correlation structure in financial returns by subtraction of the market mode. Phys. Rev. E 76 (2), 026104.

Burggraf, T., 2021. Beyond risk parity–a machine learning-based hierarchical risk parity approach on cryptocurrencies. Finance Res. Lett. 38, 101523.

Dai, Y.-H., Xie, W.-J., Jiang, Z.-Q., Jiang, G.J., Zhou, W.-X., 2016. Correlation structure and principal components in the global crude oil market. Empir. Econ. 51 (4), 1501–1519.

De Prado, M.L., 2016. Building diversified portfolios that outperform out of sample. J. Portfolio Manag. 42 (4), 59–69.

Di Matteo, T., Pozzi, F., Aste, T., 2010. The use of dynamical networks to detect the hierarchical organization of financial market sectors. Eur. Phys. J. B 73 (1), 3–11.

Götze, F., Tikhomirov, A., 2004. Rate of convergence in probability to the Marchenko-Pastur law. Bernoulli 10 (3), 503–548.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Rev. Financ. Stud. 33 (5), 2223–2273.

Haluszczynski, A., Laut, I., Modest, H., Räth, C., 2017. Linear and nonlinear market correlations: Characterizing financial crises and portfolio optimization. Phys. Rev. E 96 (6), 062315.

Han, R.-Q., Xie, W.-J., Xiong, X., Zhang, W., Zhou, W.-X., 2017. Market correlation structure changes around the great crash: A random matrix theory analysis of the Chinese stock market. Fluct. Noise Lett. 16 (02), 1750018.

Jiang, X., Chen, T., Zheng, B., 2014. Structure of local interactions in complex financial dynamics. Sci. Rep. 4 (1), 1–9.

Kruskal, J.B., 1956. On the shortest spanning subtree of a graph and the traveling salesman problem. Proc. Amer. Math. Soc. 7 (1), 48–50.

Ku, S., Lee, C., Chang, W., Song, J.W., 2020. Fractal structure in the S&P500: A correlation-based threshold network approach. Chaos Solitons Fractals 137, 109848.

Kumar, S., Deo, N., 2012. Correlation and network analysis of global financial indices. Phys. Rev. E 86 (2), 026101.

Laloux, L., Cizeau, P., Bouchaud, J.-P., Potters, M., 1999. Noise dressing of financial correlation matrices. Phys. Rev. Lett. 83 (7), 1467.

Mantegna, R.N., 1999. Hierarchical structure in financial markets. Eur. Phys. J. B 11 (1), 193–197.

Mantegna, R.N., Stanley, H.E., 1999. Introduction to Econophysics: Correlations and Complexity in Finance. Cambridge University Press.

Marčenko, V.A., Pastur, L.A., 1967. Distribution of eigenvalues for some sets of random matrices. Math. USSR-Sbornik 1 (4), 457.

Markowitz, H.M., 1968. Portfolio selection. In: Portfolio Selection. Yale University Press.

Marti, G., Nielsen, F., Bińkowski, M., Donnat, P., 2021. A review of two decades of correlations, hierarchies, networks and clustering in financial markets. Prog. Inform. Geomet. 245–274.

Mastromatteo, I., Zarinelli, E., Marsili, M., 2012. Reconstruction of financial networks for robust estimation of systemic risk. J. Stat. Mech. Theory Exp. 2012 (03), P03011.

Meng, H., Xie, W.-J., Zhou, W.-X., 2015. Club convergence of house prices: Evidence from China's ten key cities. Internat. J. Modern Phys. B 29 (24), 1550181.

Müllner, D., 2011. Modern hierarchical, agglomerative clustering algorithms. arXiv preprint arXiv:1109.2378.

Onnela, J.-P., Chakraborti, A., Kaski, K., Kertesz, J., Kanto, A., 2003. Dynamics of market correlations: Taxonomy and portfolio analysis. Phys. Rev. E 68 (5), 056110.

Pan, R.K., Sinha, S., 2007. Collective behavior of stock price movements in an emerging market. Phys. Rev. E 76 (4), 046116.

Park, J.H., Chang, W., Song, J.W., 2020. Link prediction in the Granger causality network of the global currency market. Physica A: Stat. Mech. Appl. 553, 124668.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L.A.N., Guhr, T., Stanley, H.E., 2002. Random matrix approach to cross correlations in financial data. Phys. Rev. E 65 (6), 066126.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L.A.N., Stanley, H.E., 1999. Universal and nonuniversal properties of cross correlations in financial time series. Phys. Rev. Lett. 83 (7), 1471.

Pozzi, F., Di Matteo, T., Aste, T., 2008. Centrality and peripherality in filtered graphs from dynamical financial correlations. Adv. Complex Syst. 11 (06), 927–950.

Pozzi, F., Di Matteo, T., Aste, T., 2013. Spread of risk across financial markets: Better to invest in the peripheries. Sci. Rep. 3 (1), 1–7.

Qian, E., 2011. Risk parity and diversification. J. Invest. 20 (1), 119–127.

Song, J.Y., Chang, W., Song, J.W., 2019. Cluster analysis on the structure of the cryptocurrency market via bitcoin–ethereum filtering. Physica A: Stat. Mech. Appl. 527, 121339.

Song, J.W., Ko, B., Chang, W., 2018. Analyzing systemic risk using non-linear marginal expected shortfall and its minimum spanning tree. Physica A: Stat. Mech. Appl. 491, 289–304.

Song, J.W., Ko, B., Cho, P., Chang, W., 2016. Time-varying causal network of the Korean financial system based on firm-specific risk premiums. Physica A: Stat. Mech. Appl. 458, 287–302.

Tumminello, M., Aste, T., Di Matteo, T., Mantegna, R.N., 2005. A tool for filtering information in complex systems. Proc. Natl. Acad. Sci. 102 (30), 10421–10426.

Utsugi, A., Ino, K., Oshikawa, M., 2004. Random matrix theory analysis of cross correlations in financial markets. Phys. Rev. E 70 (2), 026110.

Wang, G.-J., Xie, C., He, K., Stanley, H.E., 2017. Extreme risk spillover network: Application to financial institutions. Quant. Finance 17 (9), 1417–1433.

Yan, M., Zhao, X., 2018. Robust portfolio selection based on Gaussian rank correlation estimator. J. Phys. Conf. Ser. 1039 (1), 012021.