Robust Covariance Matrix Estimation and Portfolio Allocation: the case of non-homogeneous assets

Emmanuelle Jay[(a)], Thibault Soler[(a,c)], Jean-Philippe Ovarlez[(b)], Philippe De Peretti[(c)], Christophe Chorro[(c)]

05/05/2020

2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

[(a)] *Fideas Capital, Paris, France;* [(b)] *DEMR, ONERA, Université Paris-Saclay, Palaiseau, France;* [(c)] *Université Paris 1 Panthéon-Sorbonne, Paris, France*

Introduction

- Frequently used portfolio allocation processes require the estimation of the covariance matrix of the asset returns [1, 2, 3, 4]:
  - → The Sample Covariance Matrix (SCM) optimal under the Normal assumption is the most used estimator, but, financial time series might exhibit outliers,
  - → The field of robust estimation intends to deal with outliers [5, 6, 7],
  - → RMT helps in finding a solution for filtering noise [8, 9], but needs to be adapted to non-homogeneous and correlated time series [10].

- In [11] the authors found that considering sub-groups of homogeneous assets may allow for better performance.

- This paper focuses on assets classification methods:
  - → The Affinity Propagation (AP) method [12] that self-determines the number of classes,
  - → The Ascending Hierarchical Clustering (AHC) method that requires the number of classes or determines them using a predefined criterion.

Problem formulation

Variety Maximum (VarMax) portfolio

- The VarMax process, also called the Maximum Diversified Portfolio in [2], allocates assets by maximizing the Variety Ratio ($\mathcal{VR}$) of the portfolio.
- The $\mathcal{VR}$ quantifies the degree of diversification of a portfolio.

VarMax portfolio

Optimal weights $\mathbf{w}^*$ are the weights that maximize the Variety Ratio ($\mathcal{VR}$):

$$\mathbf{w}^* = \underset{\mathbf{w}}{\mathrm{argmax}} \frac{\mathbf{w}^T \mathbf{s}}{(\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w})^{1/2}}, \quad \text{s.t. } \mathbf{w}^T \mathbf{1}_m = 1 \text{ and } 0 \leq w_i \leq 1, \forall i \in [1, m],$$

where $\boldsymbol{\Sigma}$ is the covariance matrix and $\mathbf{s}$ such that $s_i = \sqrt{\boldsymbol{\Sigma}_{ii}}$, $i \in [1, m]$.

$\Rightarrow$ We focus on the VarMax process because it is the allocation process most sensitive to the covariance matrix of the assets returns.

$\Rightarrow$ Problem: $\boldsymbol{\Sigma}$ unknown $\rightarrow$ need to be estimated

Let $\mathbf{R} = \{\mathbf{r}_t\}_{t\in[1,N]}$ be the $m \times N$-matrix containing the N observations of the m asset returns, modelled as a K-factor model [13, 14] with an additive multivariate Elliptic Symmetric distributed noise [15, 16]. For each observation date $t \in [1, N]$, we then have:

$$\mathbf{r}_t = \mathbf{B}_t\, \mathbf{f}_t + \sqrt{\tau_t}\, \mathbf{C}^{1/2}\, \mathbf{x}_t, \tag{1}$$

- $\mathbf{r}_t$ is the m-vector of returns,
- $\mathbf{B}_t$ is the $m \times K$-matrix of coefficients that define the assets sensitivities of the K factors,
- $\mathbf{f}_t$ is the K-vector of random factors and common to the m assets,
- $\mathbf{x}_t$ is the m-vector of independent Gaussian white noise with unit variance and is non-correlated with the factors,
- $\mathbf{C}$ is called the $m \times m$ scatter matrix that is Toeplitz structured [17] and is time invariant over the period of observation,
- τ_t is an i.i.d positive random variables with expectation τ that is independent of the noise and the factors and drives the variance of the noise.

Problem formulation

Whitening process

Robust Consistent Estimation for **C** [18]

Let $\hat{\mathbf{C}}_{tyl} = \dfrac{m}{N} \displaystyle\sum_{t=1}^{N} \dfrac{\mathbf{r}_t\, \mathbf{r}_t^T}{\mathbf{r}_t^T\, \hat{\mathbf{C}}_{tyl}^{-1}\, \mathbf{r}_t}$ be the scatter matrix Tyler M-estimator of **R**.

As $m, N \to \infty$ such that $m/N \to c \in]0, \infty[$, we have $\left\| \mathcal{T}\left[\hat{\mathbf{C}}_{tyl}\right] - \mathbf{C} \right\| \xrightarrow{a.s.} 0$,

where $\mathcal{T}[\cdot]$ is the **Toeplitz rectification** operator $(\mathcal{T}[\mathbf{A}])_{ij} = \dfrac{1}{m} \displaystyle\sum_{l=i}^{m} a_{l,l-i+1}$.

A consistent estimator $\widetilde{\mathbf{C}}_{tyl}$ of the background scatter matrix **C** is therefore defined through observations **R** as $\widetilde{\mathbf{C}}_{tyl} = \mathcal{T}\left[\hat{\mathbf{C}}_{tyl}\right]$.

$\Longrightarrow$ The observations **R** can now be whitened through $\widetilde{\mathbf{C}}_{tyl}^{-1/2}\,\mathbf{R}$

Problem formulation

Whitening process

Behavior of whitened data [18]

Let $\mathbf{R}_w = \left(\mathcal{T}\left[\hat{\mathbf{C}}_{tyl}\right]\right)^{-1/2} \mathbf{R}$ be the whitened data and $\hat{\boldsymbol{\Sigma}}_{tyl}$ be the Tyler M-estimator of $\mathbf{R}_w$. As $m, N \to \infty$ such that $m/N \to c \in]0, \infty[$, if $\mathbf{R}_w$ does not contain any factor, then:

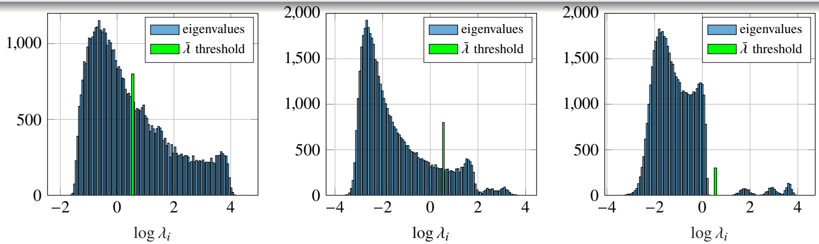
$$\left\| \hat{\boldsymbol{\Sigma}}_{tyl} - \frac{1}{N} \mathbf{X}\mathbf{X}^T \right\| \xrightarrow{a.s.} 0\,.$$

- Without factors, the spectral distribution of the whitened data scatter matrix of $\mathbf{R}_w$ follows a Marchenko-Pastur distribution [19, 20] (same spectral distribution of unobservable covariance matrix of $\mathbf{X}$) characterized by its support $\left[\left(1-\sqrt{c}\right)^2, \left(1+\sqrt{c}\right)^2\right]$,
- All eigenvalues greater than $\bar{\lambda} = \left(1+\sqrt{c}\right)^2$ can be considered as significant factors.

Estimation of K the number of factors

Let $(\lambda_k)_{k\in[1,m]}$ be the sorted eigenvalues of $\hat{\mathbf{\Sigma}}_{tyl}$, then:

$$\hat{K} = \underset{k}{\operatorname{argmax}} \left(\lambda_k > \bar{\lambda}\right)$$



Eigenvalue distributions [10]. Left: $\mathbf{R}\,\mathbf{R}^T/N$, Sample Covariance Matrix of observations. Middle: $\hat{\mathbf{C}}_{tyl}$, Tyler covariance matrix of observations. Right: $\hat{\mathbf{\Sigma}}_{tyl}$, Tyler covariance matrix of observations after whitening process. K-distributed case with shape parameter $\nu = 0.5$, $\rho = 0.8$, $m = 100$, $N = 1000$, $K = 3$.

Problem formulation

Non-homogeneous asset returns model

- The model (1) and the whitening process described above is made under the implicit assumption that the asset returns are drawn from a unique multivariate law but this assumption is unrealistic for financial time series of returns.
- To take into account the non-homogeneous asset returns, the model (1) is rewritten for the m assets splitted into $p < m$ groups. Each group is composed of $\{m_q\}_{q=1}^{p}$ assets, and composed of assets with similar distributions. It follows that:

$$\mathbf{r}_t^{(q)} = \mathbf{B}_t^{(q)} \mathbf{f}_t + \sqrt{\tau_t^{(q)}} \, \mathbf{C}_{(q)}^{1/2} \mathbf{x}_t, \tag{2}$$

$\Rightarrow$ The complete scatter matrix $\mathbf{C}$ is therefore block-constructed, block-Toeplitz, and the groups are assumed to be uncorrelated to each other.

- Under the assumption of non-homogeneous asset returns, we propose to form groups of assets before applying the whitening process.

- Two clustering methods are compared to form the groups of assets:

  ⇒ The Affinity Propagation algorithm (AP) [12] that does not require to specify the number of groups,

  ⇒ The classical Ascending Hierarchical Classification (AHC), where the number of groups p is determined arbitrarily or with Caliński-Harabasz (CH) criterion [21].

The Affinity Propagation algorithm (AP) [12]:

- an iterative partitioning method similar to the K-means, but it regroups individuals around exemplar values,
- is based on a similarity matrix $\mathbf{S}$, where $s_{i,j} = -\|\mathbf{v}_i - \mathbf{v}_j\|^2$ for $i \neq j$, and with $\mathbf{v}_i$ and $\mathbf{v}_j$ the input variables vectors of the asset i and j,
- to moderate, the number of groups p, the parameters are set to a common value using the median of pairwise similarities as in [12].

The classical Ascending Hierarchical Classification (AHC):

- is an iterative and unsupervised method,
- is based on the distances between the variables $(\mathbf{v}_i)_{i\in[1,m]}$ used to represent individuals to be grouped and seeks at each step to build the groups by aggregation,
- is used with the Euclidean distance and the Ward measure [22] to form the p groups.

Given $\mathbf{R}$ the $m \times N$-matrix of observations, and $\mathbf{R}^{(q)}$ the $m_q \times N$-matrix of observations for group (q), the whitened asset returns $\mathbf{R}_w$ are obtained through the following procedure:

- Compute the p groups using the methods described previously with $(\mathbf{v}_i)_{i\in[1,m]}$ composed of the mean μ_i, the standard deviation σ_i and of several quantiles computed from $\tilde{\mathbf{r}}_i = (\mathbf{r}_i - \mu_i \, \mathbf{1}_N) / \sigma_i$ the "standardized" returns, where $\mathbf{1}_N$ is the N-vector of ones,
- Set $\widehat{\mathbf{C}}_{tyl}^{(q)}$ the Tyler-M estimate of $\mathbf{R}^{(q)}$,
- Set $\widetilde{\mathbf{C}}_{tyl}^{(q)} = \mathcal{T}\left(\widehat{\mathbf{C}}_{tyl}^{(q)}\right)$, the Toeplitz rectification matrix built from $\widehat{\mathbf{C}}_{tyl}^{(q)}$ for the Toeplitz operator $\mathcal{T}$,
- Set $\mathbf{R}_w^{(q)} = \left(\widetilde{\mathbf{C}}_{tyl}^{(q)}\right)^{-1/2} \mathbf{R}^{(q)}$, the $m_q \times N$ matrix of the whitened observations of group q,

Proposed methodology

Detailed whitening procedure

Finally, the de-noised covariance matrix estimate $\widehat{\boldsymbol{\Sigma}}_w$ is obtained as follows:

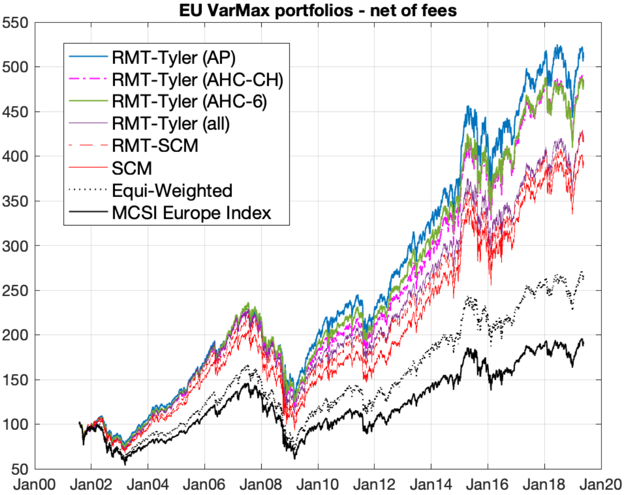
- Set $\widehat{\boldsymbol{\Sigma}}_{tyl}$ as the Tyler-M estimate of $\mathbf{R}_w$, where $\mathbf{R}_w = \left[\mathbf{R}_w^{(1)\,T} \ldots \mathbf{R}_w^{(p)\,T}\right]^T$ of size $m \times N$,
- Set $\widehat{\boldsymbol{\Sigma}}_{tyl}^{clip} = \mathbf{U}\,\boldsymbol{\Lambda}^{clip}\,\mathbf{U}^T$ where $\mathbf{U}$ is the $m \times m$ eigenvectors matrix and $\boldsymbol{\Lambda}^{clip}$ is the $m \times m$ diagonal matrix of the eigenvalues $(\lambda_k^{clip})_{k\in[1,m]}$ corrected using the Eigenvalue clipping method [23]:

$$\lambda_k^{clip} = \begin{cases} \lambda_k, & \text{if } \lambda_k \geq \left(1+\sqrt{c}\right)^2 \\ \dfrac{1}{m-K}\left(\displaystyle\sum_{k=1}^{m}\lambda_k - \sum_{k=1}^{K}\lambda_k\right), & \text{otherwise} \end{cases}$$

- Finally, $\widehat{\boldsymbol{\Sigma}}_w = \left(\widetilde{\mathbf{C}}_{tyl}^{1/2}\right)\widehat{\boldsymbol{\Sigma}}_{tyl}^{clip}\left(\widetilde{\mathbf{C}}_{tyl}^{1/2}\right)^T$.

- Two investment universes are tested:
  - $\Rightarrow$ European equity indices ($m = 43$): countries, sub-sectors and factors.
  - $\Rightarrow$ US equity indices ($m = 30$): sub-sectors and factors.
- Optimization settings:
  - $\Rightarrow$ daily closing prices from July 27th, 2000 to May 20th, 2019,
  - $\Rightarrow$ the covariance matrix of the assets is estimated using the past daily returns ($N = 260$),
  - $\Rightarrow$ the portfolio weights are computed every four weeks and kept constant for the next four-weeks period.
- Clustering method settings:
  - $\Rightarrow$ the quantiles used are q_θ and $q_{1-\theta}$ with $\theta \in \{1\%, 2.5\%, 5\%, 10\%, 15\%, 25\%, 50\%\}$,
  - $\Rightarrow$ for AHC method, $p = 6$ ("AHC-6") as in [11] or set according to the CH criterion ("AHC-CH").
- The portfolio performances are net of transaction fees (0.07%) to take into account the portfolio turnover.

Application
EU portfolio results

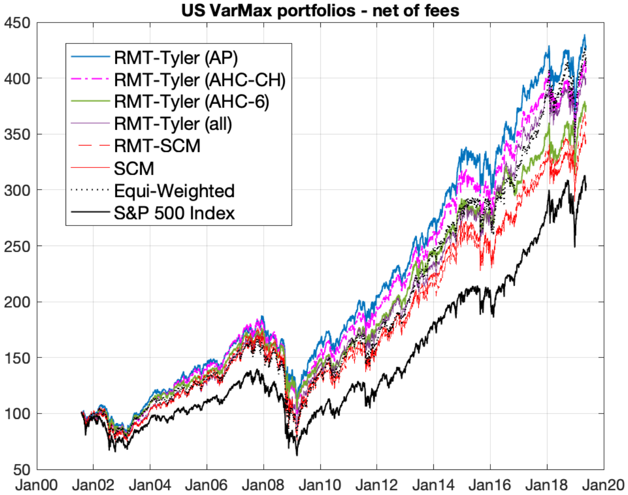


EU VarMax portfolios' wealth with 0.07% of fees from July 2001 to May 2019.

EU VarMax Portfolios	Ann. Ret.	Ann. Vol.	Ratio Ret/Vol	Max Drawdown	$\mathcal{VR}$ (avg)
RMT-Tyler (AP)	9.87%	12.14%	0.81	45.37%	1.46
RMT-Tyler (AHC-6)	9.65%	12.03%	0.80	46.84%	1.57
RMT-Tyler (AHC-CH)	9.58%	12.45%	0.77	48.16%	1.51
RMT-Tyler (all)	8.90%	13.16%	0.68	51.18%	1.44
RMT-SCM	8.94%	13.79%	0.65	54.15%	1.27
SCM	8.56%	13.68%	0.63	54.45%	1.38
Equi-Weighted	6.60%	15.37%	0.43	57.82%	1.19
MSCI Europe	*4.71%*	*14.87%*	*0.32*	*58.54%*	

Performance numbers for the Europe (EU) VarMax portfolios with 0.07% of fees from July 2001 to May 2019.

Application
US portfolio results



US VarMax portfolios' wealth with 0.07% of fees from July 2001 to May 2019.

Application

US portfolio results

US VarMax Portfolios	Ann. Ret.	Ann. Vol.	Ratio Ret/Vol	Max Drawdown	$\mathcal{VR}$ (avg)
RMT-Tyler (AP)	8.76%	11.11%	0.79	42.82%	1.51
RMT-Tyler (AHC-CH)	8.57%	11.53%	0.74	46.57%	1.55
RMT-Tyler (AHC-6)	7.98%	10.79%	0.74	41.50%	1.52
RMT-Tyler (all)	8.49%	12.09%	0.70	49.27%	1.53
Equi-Weighted	8.92%	13.83%	0.65	53.70%	1.25
RMT-SCM	8.03%	13.13%	0.61	56.53%	1.34
SCM	7.80%	13.27%	0.59	55.47%	1.46
S&P 500	*7.21%*	*14.18%*	*0.51*	*55.71%*	

Performance numbers for the US VarMax portfolios with 0.07% of fees from July 2001 to May 2019.

Conclusion

- Asset returns have been modelled as a multi-factor model embedded in a correlated elliptical and symmetric noise by considering that the asset returns are non-homogeneous in law which is more realistic,
- Given this model setup, we question the ability of classification methods (AP algorithm and AHC) to improve whitening process based on the Tyler M-estimator and the RMT,
- Our methodology has been tested on the Maximum Variety portfolio optimization problem and proves the superiority of the AP algorithm in producing higher performances for both EU and US universes.

References I

[1] H. M. Markowitz. “Portfolio Selection”. In: *Journal of Finance* 7.1 (1952), pp. 77–91.

[2] Y. Choueifaty and Y. Coignard. “Toward maximum diversification”. In: *Journal of Portfolio Management* 35.1 (2008), pp. 40–51.

[3] S. Maillard, T. Roncalli, and J. Teiletche. “The properties of equally weighted risk contributions portfolios”. In: *Journal of Portfolio Management* 36 (2010), pp. 60–70.

[4] R. Clarke, H. De Silva, and S. Thorley. “Minimum variance, maximum diversification, and risk parity: an analytic perspective”. In: *Journal of Portfolio Management* (2012).

[5] R. A. Maronna. “Robust M-Estimators of Multivariate Location and Scatter”. In: *Annals of Statistics* 4.1 (1976), pp. 51–67.

[6] D. E. Tyler. “A distribution-free M-estimator of multivariate scatter”. In: *The annals of Statistics* 15.1 (1987), pp. 234–251.

References II

[7] J. Vinogradova, R. Couillet, and W. Hachem. "Estimation of Toeplitz Covariance Matrices in Large Dimensional Regime With Application to Source Detection". In: *IEEE Trans. on Signal Processing* 63.18 (2015), pp. 4903–4913. ISSN: 1053-587X.

[8] L. Laloux et al. "Noise Dressing of Financial Correlation Matrices". In: *Physycal Review Letters* 83.1468 (1999).

[9] M. Potters, J. P. Bouchaud, and L. Laloux. "Financial applications of Random Matrix Theory: old laces and new pieces". In: *Acta Physica Polonica B* 36.9 (2005).

[10] E. Jay et al. "Improving Portfolios Global Performance with Robust Covariance Matrix Estimation: Application to the Maximum Variety Portfolio". In: *26th EUSIPCO*. 2018.

[11] E. Jay et al. "Improving portfolios global performance using a cleaned and robust covariance matrix estimate". In: *Soft Computing* (2020).

[12] B. J. Frey and D. Dueck. "Clustering by Passing Messages Between Data Points". In: *Science* 315.5814 (2007), pp. 972–976.

References III

[13] E. Jay et al. “Multi-factor Models: examining the potential of signal processing techniques”. In: *IEEE Signal Processing Magazine* 28.5 (2011).

[14] S. Darolles, C. Gouriéroux, and E. Jay. “Robust portfolio allocation with risk contribution restrictions”. In: *Forum GI - Paris*. 2013.

[15] D. Kelker. “Distribution theory of spherical distributions and a location-scale parameter generalization”. In: *Sankhyā: The Indian Journal of Statistics, Series A* 32.4 (1970), pp. 419–430. ISSN: 0581-572X.

[16] E. Ollila et al. “Complex Elliptically Symmetric Distributions: Survey, New Results and Applications”. In: *IEEE Trans. on Signal Processing* 60.11 (2012), pp. 5597–5625. ISSN: 1053-587X.

[17] R. M. Gray. “Toeplitz and Circulant Matrices: A Review”. In: *Foundations and Trends® in Communications and Information Theory* 2.3 (2006), pp. 155–239.

[18] E. Terreaux, J. P. Ovarlez, and F. Pascal. “New model order selection in large dimension regime for Complex Elliptically Symmetric noise”. In: *25th EUSIPCO*. 2017, pp. 1090–1094.

[19] V. A. Marchenko and L. A. Pastur. “Distribution of eigenvalues for some sets of random matrices”. In: *Matematicheskii Sbornik* (1967).

References IV

[20] T. Zhang, C. Xiuyuan, and A. Singer. "Marcenko-Pastur Law for Tyler's and Maronna's M-estimator". In: *Journal of Multivariate Analysis* 149 (2016), pp. 114–123.

[21] T. Caliński and J. Harabasz. "A dendrite method for cluster analysis". In: *Communications in Statistics* 3.1 (1974), pp. 1–27.

[22] J. H. Jr. Ward. "Hierarchical grouping to optimize an objective function". In: *Journal of the American Statistical Association* 58 (1963), pp. 236–244.

[23] L. Laloux et al. "Random Matrix Theory and financial correlations". In: *International Journal of Theoretical and Applied Finance* 3.03 (2000), pp. 391–397.