# Improving Portfolios Global Performance with Robust Covariance Matrix Estimation: Application to the Maximum Variety Portfolio

Emmanuelle Jay[(a)], Eugénie Terreaux[(b)], Jean-Philippe Ovarlez[(b,c)] and Frédéric Pascal[(b)]

[(a)] *Fideas Capital, F-75001 Paris, France;* [(b)] *CentraleSupélec, Université Paris-Saclay, F91190 Gif-sur-Yvette, France;* [(c)] *ONERA, Université Paris-Saclay, F-91123, Palaiseau, France*

## Problem Setting

- Frequently used portfolio allocation processes require the estimation of the covariance matrix of the assets returns
  *(e.g. Global Minimum Variance [Maillard 10, Clarke 12], Maximum Variety [Fideas Capital] or Most Diversified Portfolio [Choueifaty 08], Mean-Variance [Markovitz 52], etc.)*
  → The Sample Covariance Matrix (SCM) - optimal under the Normal assumption - is the mostly used estimator, but, financial time series of returns might exhibit outliers,
  → The field of robust estimation intends to deal with outliers ([*Maronna 76, Tyler 87*]),
  → Hybrid robust shrinkage covariance matrix estimates have also been proposed building estimators upon Tyler's robust M-estimator ([*Chen 11, Ollila 14, Pascal 14*]),
  → Recent works based on Random Matrix Theory (RMT) have also considered robust estimation in the large dimensional regime ([*Yang 15*]).
- A way to mitigate covariance matrix estimation errors is to identify the most informative asset part and then to filter the noisy part of the data
  → Standard statistical methods like the principal component analysis may fail in distinguishing informative factors from the noisy ones,
  → RMT helps in finding a solution for filtering noise, even though the single market factor still prevails in the described cleaning method that is not completely satisfactory as they implicitly assumes homogeneous and uncorrelated series ([*Laloux 99 and 00, Potters 05*]),
  ⇒ To fill this gap, the most up-to-date RMT-based model order selection [*Vinogradova 13, Terreaux 17*] methods used in Signal Processing can be applied in estimating the number of uncorrelated statistical factors embedded in a given multi-factor model.

### We propose in this paper...

... **a new process for estimating and denoising covariance matrix that leads to improved global portfolio performances *(reduced Draw-Down, increased Sharpe Ratio, etc.)*.**

→ Asset returns are modelled as a multi-factor model ([*Jay 11, Darolles 13*]),
→ An up-to-date Model Order Selection method is used to estimate the number of factors,
→ It can be easily applied in many Signal Processing applications like in radar and sonar (Direction of Arrival, Source Localization, Space Time Adaptive Processing, Date of Arrival, Spectral Analysis (AR, ARMA), etc), Hyperspectral images (Unmixing).

## Assets returns Model

Let $\{\mathbf{r}_t\}_{t\in[1,N]}$ be $N$ observations of the $m$ assets returns, modelled as a $K$-factor model. For each observation date $t$, we then have:

$$\mathbf{r}_t = \sum_{k=1}^{K} \mathbf{f}_{t,k}\,\boldsymbol{\beta}_k + \sqrt{\tau_t}\,\mathbf{C}^{1/2}\,\mathbf{n}_t\,, \quad t\in[1,N]\,,$$

or, written more compactly: $\mathbf{R} = \mathbf{B}\,\mathbf{F} + \mathbf{C}^{1/2}\,\mathbf{N}\,\mathbf{T}^{1/2}$, where

- $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N] \in \mathbb{R}^{m\times N}$ are the observations,
- $\mathbf{B} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K) \in \mathbb{R}^{m\times K}$ is an unknown mixing matrix of coefficients (or beta) that define the proportion of the $K$ factors in each asset,
- $\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_N) \in \mathbb{R}^{K\times N}$ is an unknown matrix of the $K$ common returns,
- $\mathbf{T} = \mathrm{diag}(\tau_1, \dots, \tau_N) \in \mathbb{R}^{+N\times N}$ is a diagonal matrix unknown containing random texture,
- $\mathbf{N} \in \mathbb{R}^{m\times N}$ is a white Gaussian noise ($E\left[\mathbf{n}_t^T\,\mathbf{n}_t\right] = 1$), independent of the $K$ factors,
- $\mathbf{C} \in \mathbb{R}^{m\times m}$ is an unknown **Toeplitz** scatter matrix ($\mathrm{Tr}(\mathbf{C}) = m$).

## Theoretical Results [*Terreaux 17*]

### Robust Consistent Estimation for C

Let $\hat{\mathbf{M}}_{FP} = \dfrac{m}{N}\sum_{t=1}^{N}\dfrac{\mathbf{r}_t\,\mathbf{r}_t^T}{\mathbf{r}_t^T\,\hat{\mathbf{M}}_{FP}^{-1}\,\mathbf{r}_t}$ be the scatter matrix Tyler M-estimator of $\mathbf{R}$.
As $m, N \to \infty$ such that $m/N \to c \in ]0,\infty[$, we have

$$\left\|\mathcal{T}\left[\hat{\mathbf{M}}_{FP}\right] - \mathbf{C}\right\| \xrightarrow{a.s.} 0,$$

where $\mathcal{T}[\cdot]$ is the **Toeplitz rectification** operator: $(\mathcal{T}[\mathbf{X}])_{ij} = \dfrac{1}{m}\sum_{k=1}^{m}\mathbf{X}_{k,k+|i-j|}$.

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

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

### Behavior of whitened data

Let $\mathbf{R}_w = \left(\mathcal{T}\left[\hat{\mathbf{M}}_{FP}\right]\right)^{-1/2}\mathbf{R}$ be the whitened data and $\hat{\mathbf{W}}_{FP}$ 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:

$$\left\|\hat{\mathbf{W}}_{FP} - \frac{1}{N}\mathbf{N}\,\mathbf{N}^T\right\| \xrightarrow{a.s.} 0\,.$$

- Without factors, the spectral distribution of the whitened data scatter matrix of $\mathbf{R}_w$ follows a Marchenko-Pastur distribution (same spectral distribution of unobservable covariance matrix of $\mathbf{N}$) characterized by its support $\left[(1-\sqrt{c})^2, (1+\sqrt{c})^2\right]$,
- All eigenvalues greater than $\bar{\lambda} = (1+\sqrt{c})^2$ 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 $\hat{\mathbf{W}}_{FP}$, then: $\hat{K} = \underset{k}{\mathrm{argmax}}\left(\lambda_k > \bar{\lambda}\right)$.

### Illustration: estimating the correct number of factors

Estimating $K$ is really a challenging problem for many applications where informative signal is embedded in correlated noise. Below, we show how our process allows to detect the $K = 3$ sources embedded in non-Gaussian and strongly correlated noise.

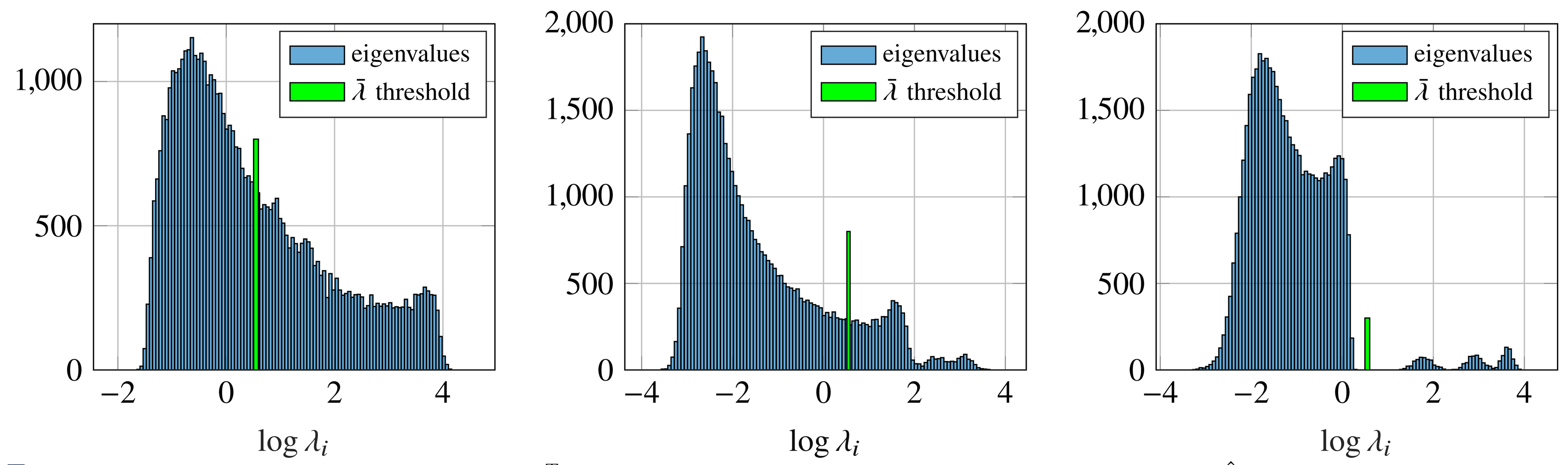


Figure 1: Eigenvalue distributions. Left: $\mathbf{R}\,\mathbf{R}^T/N$, Sample Covariance Matrix of observations. Middle: $\hat{\mathbf{M}}_{FP}$, Tyler covariance matrix of observations. Right: $\hat{\mathbf{W}}_{FP}$, Tyler covariance matrix of observations after whitening process. K-distributed case with shape parameter $\nu = 0.5$, $\rho = 0.8$, $m = 100$, $N = 1000$, $K = 3$.

## Maximum Variety Portfolio

One way to quantify the degree of diversification of a portfolio invested in $m$ assets with proportions $\mathbf{w} = [w_1, \dots, w_m]^T$ is to maximize the Variety Ratio of the portfolio:

$$\mathbf{w}_{vr}^* = \underset{\mathbf{w}}{\mathrm{argmax}}\frac{\mathbf{w}^T\,\mathbf{s}}{\left(\mathbf{w}^T\,\boldsymbol{\Sigma}\,\mathbf{w}\right)^{1/2}},$$

where $\boldsymbol{\Sigma} = E\left[\mathbf{R}\,\mathbf{R}^T\right]$ is the $m\times m$ covariance matrix of the $m$ assets returns $\mathbf{R}$ and where $\mathbf{s}$ is the $m$-vector of the square roots of the diagonal element of $\boldsymbol{\Sigma}$, *ie* $s_i = \sqrt{\boldsymbol{\Sigma}_{ii}}$, $i\in[1,m]$, representing the standard deviation of the returns of the $m$ assets.

## Application

The investment universe consists of $m = 40$ baskets of European equity stocks representing twenty-one industry subsectors (e.g. transportation, materials, etc.), thirteen countries (e.g. Sweden, France, etc.) and six factor-based indices (e.g. momentum, quality, growth, etc.)

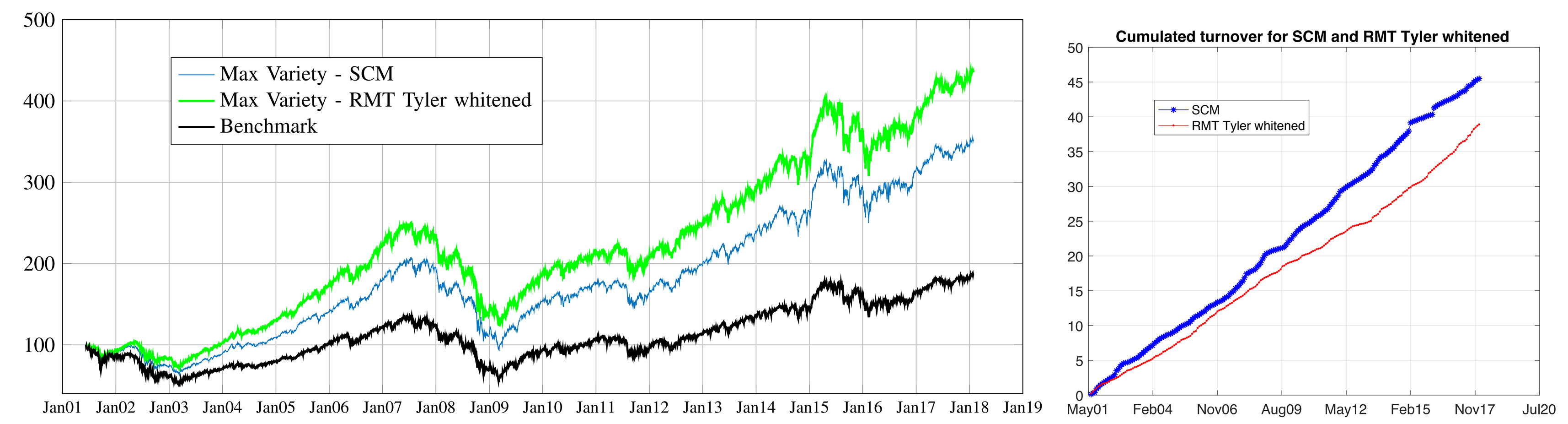


Figure 2: Left: portfolios wealth starting at 100 at the first period. Right: cumulative sum of absolute weight changes (turnover) between the consecutive periods. Performance are compared to the STOXX® Europe 600 Index performance composed of 600 large, mid and small equity stocks across 17 countries of the European regions.

Table 1: Some performance numbers.

| Variety Max Portfolios | Annualised Return | Annualised Volatility | Ratio (Return / Volatility) | Maximum Drawdown |
|---|---|---|---|---|
| RMT Tyler Whitened | 9.71% | 12.9% | 0.75 | 50.41% |
| SCM | 8.51% | 13.80% | 0.62 | 55.02% |
| Benchmark | 4.92% | 15.19% | 0.32 | 58.36% |

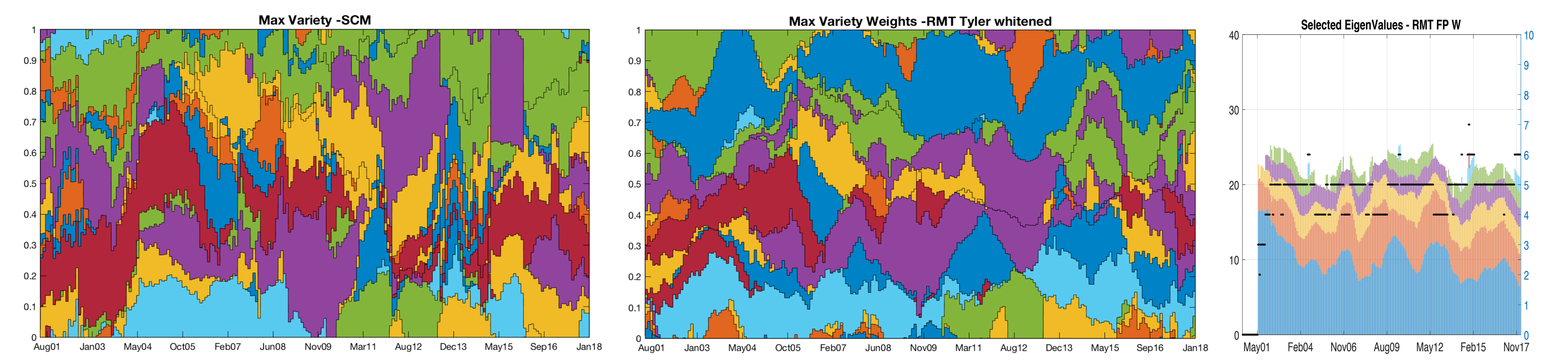


Figure 3: Left - Middle: dynamic weights as a stacked area chart. Right: values of selected eigenvalues (left) and their number (right).

## Conclusion

- Asset returns have been modelled as a multi-factor model embedded in a correlated elliptical and symmetric noise, allowing to account for non-Gaussian and non correlated noise,
- Given this model setup, the most informative assets have been separated from the noise subspace using a '*Toeplitzified*' robust and consistent Tyler-M estimator and the Random Matrix Theory applied on the whitened covariance matrix estimate,
- As an illustration, applied to the Maximum Variety Portfolios, our process leads to improved performance with respect to a classical approach.

**Conf. EUSIPCO 2018** | *This research was both conducted within the "Construction of factorial indexes and allocation" under the aegis of the Europlace Institut of Finance, a joint initiative with Fideas Capital and was partly funded by the French DGA* | **Roma, 5 Sept. 2018**