



Covariance matrix estimation for robust portfolio allocation

Ahmad W. Bitar, Nathan de Carvalho, Valentin Gatignol

► To cite this version:

Ahmad W. Bitar, Nathan de Carvalho, Valentin Gatignol. Covariance matrix estimation for robust portfolio allocation. CentraleSupélec; Université de Technologie de Troyes; Université Paris Cité; Engie Global Markets; Qube Research and Technologies. 2023. hal-04046454v2

HAL Id: hal-04046454

<https://centralesupelec.hal.science/hal-04046454v2>

Submitted on 26 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Covariance matrix estimation for robust portfolio allocation

Ahmad W. Bitar^{*1}, Nathan De Carvalho^{*2}, Valentin Gatignol^{*3}

^{*} CentraleSupélec, Université Paris-Saclay, F-91190 Gif-sur-Yvette, France

¹ Université de Technologie de Troyes, LIST3N-M2S, Troyes, France

²Engie Global Markets & Université Paris Cité and Sorbonne Université, LPSM, Paris, France

³Qube Research & Technologies, Paris, France

Abstract

In this technical report¹, we aim to combine different portfolio allocation techniques with covariance matrix estimators to meet two types of clients' requirements: client A who wants to invest money wisely, not taking too much risk, and not willing to pay too much in rebalancing fees; and client B who wants to make money quickly, benefit from market's short-term volatility, and ready to pay rebalancing fees. Four portfolio techniques are considered (mean-variance, robust portfolio, minimum-variance, and equi-risk budgeting), and four covariance estimators are applied (sample covariance, ordinary least squares (OLS) covariance, cross-validated eigenvalue shrinkage covariance, and eigenvalue clipping). Some comparisons between the covariance estimators in terms of eigenvalue stability and four metrics (i.e. expected risk, gross leverage, Sharpe ratio and effective diversification) exhibit the superiority of the eigenvalue clipping covariance estimator. The experiments on the Russel1000 dataset show that the minimum-variance with eigenvalue clipping is the model suitable for client A, whereas robust portfolio with eigenvalue clipping is the one suitable for client B.

Index Terms

Robust portfolio, minimum-variance, eigenvalue clipping, OLS covariance.

I. INTRODUCTION

Financial crises have large deleterious effects on economic activity and have been well marked in the history since the late 1990s. In this regard, economists have become more engaged with the study of financial crises since their probability of occurrence has increased dramatically especially with the last 20 years [1]. Modelling financial market risk has thus become primordial and is the reason why the focus of portfolio allocation has shifted from the return (μ) side of the (μ, σ) coin to the risk (σ) side. The expected return of a portfolio can be expressed as

$$\mu = \boldsymbol{\mu}^T \mathbf{w},$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\mathbf{w} = [w_1, \dots, w_p]^T \in \mathbb{R}^p$ are the vectors of asset expected returns and weights, respectively. The risk of a portfolio can be expressed as

$$\sigma = \sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}},$$

where $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$ is the unknown covariance matrix of asset returns that has to be efficiently estimated especially in large dimensions.

The major breakthrough came in 1952 with the publication of Harry Markowitz's theory of portfolio selection and which laid the groundwork for modern portfolio construction theory with the so-called mean-variance optimization [2]. The latter aims to select the optimal portfolio by achieving a compromise between unknown portfolio expected return (μ) and unknown portfolio risk (σ). More precisely, the mean-variance optimization (with soft constraint to risk aversion) is defined as

$$\begin{aligned} \max_{\mathbf{w} \in \mathbb{R}^p} \left\{ \boldsymbol{\mu}^T \mathbf{w} - \frac{\lambda}{2} \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} - \alpha \|\mathbf{w} - \mathbf{w}_{previous}\|_{1 \text{ or } 2} \right\} \\ \text{s.t.} \quad \mathbf{1}_p^T \mathbf{w} = 1, \\ w_j \geq 0, \quad j \in [1, p], \end{aligned} \quad (1)$$

where λ is the risk aversion intensity of the investor, $\mathbf{1}_p$ is a vector of ones of size p , α is the aversion parameter to turnover, and $\mathbf{w}_{previous}$ is obtained from previous optimization.

The optimal portfolio selected by problem (1) is not robust against small changes in inputs especially in assets mean returns and can lead to large changes in portfolio allocation. In addition, assets with positive estimation error will usually be over-weighted, whereas assets with negative estimation error will be under-estimated [3]. In [4] and [5], the authors proposed to further constraining the mean-variance optimization to keep the optimal portfolio within feasible bounds. However, the selection of the constraints is highly subjective and the optimal portfolio is quite sensitive to such selection [3]. To reduce statistical errors in the mean returns, the latter can be shrinked towards a sensible value. For example, in [6] and [7], the

¹This report overviews in detail our final project in the course portfolio allocation presented at CentraleSupélec on March 19, 2021. We hereby certify that this is entirely our own developed work unless otherwise stated.

authors proposed to shrink the mean returns towards an equal returns and market-implied returns, respectively. These shrinkage approaches help in increasing the robustness but are quite sensitive to the selection of the shrinkage parameter and can affect the optimal portfolio allocation very badly [3]. In order to take into account the uncertainty in the return estimation, it would be interesting to reformulate the mean-variance optimization problem by taking into consideration the estimation errors which may help to reduce the sensitivity of the portfolios as the estimation of errors increases. This modification has lead to the so-called robust portfolio optimization [3], [8], [9], [10], [11], [12]. More precisely, the mean-variance portfolio (with soft constraint to risk aversion) can be modified as

$$\begin{aligned} \max_{\mathbf{w} \in \mathbb{R}^p} \left\{ \min_{\boldsymbol{\mu} \in \mathbb{R}^p} \{ \boldsymbol{\mu}^T \mathbf{w} \} - \frac{\lambda}{2} \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} - \theta \|\mathbf{w} - \mathbf{w}_{previous}\|_{1 \text{ or } 2} \right\} \\ \text{s.t.} \quad (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^T \boldsymbol{\Omega}^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) < \kappa^2 \\ \mathbf{1}_p^T \mathbf{w} = 1, \\ w_j \geq 0, \quad j \in [1, p], \end{aligned} \quad (2)$$

where $\hat{\boldsymbol{\mu}}$ is the estimated mean returns, κ is the aversion parameter to error, $\boldsymbol{\Omega}$ is the covariance matrix of errors, and θ is the aversion parameter to turnover. Problem (2) implies that one needs to maximize the portfolio returns in the worst case scenario which is defined by letting the estimated mean returns to travel in an uncertainty box about the true mean returns (that is, $(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^T \boldsymbol{\Omega}^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) < \kappa^2$).

We have

$$\begin{aligned} \min_{\boldsymbol{\mu} \in \mathbb{R}^p} \{ \boldsymbol{\mu}^T \mathbf{w} \} &\implies \max_{\boldsymbol{\mu} \in \mathbb{R}^p} \{ \hat{\boldsymbol{\mu}}^T \mathbf{w} - \boldsymbol{\mu}^T \mathbf{w} \} \\ \text{s.t.} \quad (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^T \boldsymbol{\Omega}^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) &< \kappa^2. \end{aligned}$$

The Lagrangian function (with a Lagrangian multiplier Φ) associated with $\max_{\boldsymbol{\mu} \in \mathbb{R}^p} \{ \hat{\boldsymbol{\mu}}^T \mathbf{w} - \boldsymbol{\mu}^T \mathbf{w} \}$ can be defined as

$$\mathcal{L}(\boldsymbol{\mu}, \Phi) = \hat{\boldsymbol{\mu}}^T \mathbf{w} - \boldsymbol{\mu}^T \mathbf{w} - \Phi \left((\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^T \boldsymbol{\Omega}^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) - \kappa^2 \right).$$

Equating to zero both partial derivatives of $\mathcal{L}(\boldsymbol{\mu}, \Phi)$ relative to $\boldsymbol{\mu}$ and Φ and then solving for $\boldsymbol{\mu}$ leads to the solution [3], [8]

$$\boldsymbol{\mu} = \hat{\boldsymbol{\mu}} - \sqrt{\frac{\kappa^2}{\mathbf{w}^T \boldsymbol{\Omega} \mathbf{w}}} \boldsymbol{\Omega} \mathbf{w}.$$

At the end, one obtains the following robust portfolio optimization [3], [8], [9], [10], [11], [12]:

$$\begin{aligned} \max_{\mathbf{w} \in \mathbb{R}^p} \left\{ \hat{\boldsymbol{\mu}}^T \mathbf{w} - \kappa \sqrt{\mathbf{w}^T \boldsymbol{\Omega} \mathbf{w}} - \frac{\lambda}{2} \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} - \theta \|\mathbf{w} - \mathbf{w}_{previous}\|_{1 \text{ or } 2} \right\} \\ \text{s.t.} \quad \mathbf{1}_p^T \mathbf{w} = 1, \\ w_j \geq 0, \quad j \in [1, p]. \end{aligned} \quad (3)$$

As we can observe, if κ is sufficiently small, the term $\kappa \sqrt{\mathbf{w}^T \boldsymbol{\Omega} \mathbf{w}}$ will vanish, and thus, problem (3) reduces to problem (1). On the other hand, if κ is large, the optimal solution to problem (3) can deviate significantly from the solution of problem (1) [3]. It is important to note that problem (3) can no longer be solved using quadratic programming due to the presence of the square root. It must be solved using either second-order cone programming or an optimizer that can handle general convex expressions. These optimization algorithms are readily available in many standard statistical software packages and run almost as fast as traditional quadratic algorithms for practical portfolio optimization applications [3].

However, the aforementioned optimization (returns-oriented) strategies depend on the expected return and the covariance matrix $\boldsymbol{\Sigma}$, where both of them are completely unknown and have to be carefully estimated. As was discussed in [13], [14], several other approaches are seen to avoid any or all of the mean returns, variance and covariances. For example, the minimum-variance and equi-risk budgeting portfolios [15], [16] are kinds of risk-oriented portfolio strategies that aim to focus solely on the covariance matrix $\boldsymbol{\Sigma}$. More precisely, the minimum-variance portfolio (with soft constraint to risk aversion) can be defined as

$$\begin{aligned} \min_{\mathbf{w} \in \mathbb{R}^p} \{ \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} + \beta \|\mathbf{w} - \mathbf{w}_{previous}\|_{1 \text{ or } 2} \} \\ \text{s.t.} \quad \mathbf{1}_p^T \mathbf{w} = 1, \\ w_j \geq \text{min_short}, \quad j \in [1, p], \end{aligned} \quad (4)$$

where β is the aversion parameter to turnover. The equi-risk budgeting portfolio is defined as

$$\begin{aligned} & \min_{\mathbf{w} \in \mathbb{R}^p} \{ \mathbf{w}^T \Sigma \mathbf{w} \} \\ \text{s.t.} \quad & \mathbf{1}_p^T \log \mathbf{w} = c, \\ & w_j \geq 0, \quad j \in [1, p], \end{aligned} \quad (5)$$

where c is a constant.

The minimum-variance portfolio is mean-variance efficient and achieves the highest Sharpe ratio if the asset mean returns are all equal. In addition, it is known to usually give low-diversified but performing portfolios reinforcing the low-volatility anomaly concept [3]. However, the equi-risk budgeting portfolio is mean-variance efficient with the highest Sharpe ratio portfolio if the asset mean returns are proportional to the asset volatilities in addition to that the return correlations are all equal [3].

In this technical report, we aim to combine the aforementioned portfolio optimization techniques (mean-variance, robust portfolio, minimum-variance, and equi-risk budgeting) with several covariance matrix estimators (sample covariance matrix $\hat{\Sigma}_{SCM}$, OLS covariance matrix $\hat{\Sigma}_{OLS}$ [17], [18], [19], cross-validated eigenvalue shrinkage covariance matrix $\hat{\Sigma}_{CV}$ [20], and eigenvalue clipping $\hat{\Sigma}_{Clipping}$ [21]) to meet two types of clients' requirements:

- client A: (1) wants to invest money wisely; (2) not taking too much risk; and (3) not willing to pay too much in rebalancing fees;
- client B: (1) wants to make money quickly; (2) benefit from market's short-term volatility; and (3) ready to pay rebalancing fees.

Based on some experiments on the Russel1000 dataset in section III, our results show that the robust portfolio with eigenvalue clipping is the model suitable for client B, whereas minimum-variance with eigenvalue clipping is the one suitable for client A.

The remainder of this technical report is organized as follows. Some of the aforementioned covariance estimators are briefly described in section II. The experiments based on the eigenvalue stability and some metrics are presented in section III. Finally, section IV gives concluding remarks and some directions for future work.

Notations: throughout this technical report, we depict vectors in lowercase boldface letters and matrices in uppercase boldface letters. The notation $(\cdot)^T$, $|\cdot|$, and $\text{Tr}(\cdot)$ stand for the transpose, determinant, and trace of a matrix, respectively.

II. COVARIANCE MATRIX ESTIMATORS

The estimation of the covariance matrix is of great interest in finance as it is fundamentally the building block to derive optimal investment strategies. It is usually done via a collection of samples that can be called "secondary data". The covariance matrix estimators that are going to be evaluated in this paper are the following:

- Sample Covariance Matrix (SCM) estimator, $\hat{\Sigma}_{SCM}$;
- Ordinary Least Squares (OLS) Covariance estimator $\hat{\Sigma}_{OLS}$ [17], [18], [19];
- Eigenvalue clipping estimator, $\hat{\Sigma}_{Clipping}$ [21];
- Cross-validated (CV) eigenvalue shrinkage covariance estimator, $\hat{\Sigma}_{CV}$ [20].

In what follows, we only present in detail three covariance matrix estimators: $\hat{\Sigma}_{SCM}$, $\hat{\Sigma}_{OLS}$, and $\hat{\Sigma}_{Clipping}$. For the details about $\hat{\Sigma}_{CV}$, please refer to [20].

A. The SCM estimator: $\hat{\Sigma}_{SCM}$

We already know that in practice, it is rare to perfectly know the distribution of the data and the Gaussian model assumption is the commonly widely hypothesis used in several applications. Assume we observe n independent and identically distributed (i.i.d) p -random vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$, distributed according to a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and unknown covariance matrix Σ . That is, $\mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ with $i \in [1, n]$.

The likelihood function of the matrix $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix} \in \mathbb{R}^{n \times p}$ is defined as

$$\begin{aligned} \mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_n; \boldsymbol{\mu}, \Sigma) = \\ \prod_{i=1}^n (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right). \end{aligned}$$

The log-likelihood function of the matrix \mathbf{X} is obtained as

$$\log(\mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_n; \boldsymbol{\mu}, \boldsymbol{\Sigma})) = \frac{-np}{2} \log(2\pi) + \frac{n}{2} \log|\boldsymbol{\Sigma}^{-1}| - \frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}).$$

By maximizing the log-likelihood w.r.t. $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, one obtains the following estimators

$$\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \frac{1}{n} \mathbf{X}^T \mathbf{1}_n,$$

and

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_{SCM} &= \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T - \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T = \frac{1}{n} \mathbf{X}^T \mathbf{X} - \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^T \\ &= \frac{1}{n} \mathbf{X}^T \mathbf{X} - \frac{1}{n^2} \mathbf{X}^T (\mathbf{1}_n \mathbf{1}_n^T) \mathbf{X} \\ &= \mathbf{X}^T \left(\frac{1}{n} \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right) \right) \mathbf{X} \\ &= \frac{1}{n} \mathbf{X}^T \mathbf{H}_n \mathbf{X}, \end{aligned}$$

where $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Sigma}}_{SCM}$, \mathbf{I}_n , $\mathbf{H}_n = (\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T) \in \mathbb{R}^{n \times n}$, and $\mathbf{1}_n$ are the sample mean, (biased) Sample Covariance matrix (SCM), the identity matrix of size $n \times n$, the centering matrix, and a vector of ones of size n , respectively.

(i) The matrix $\mathbf{H}_n^2 = \mathbf{H}_n$:

$$\begin{aligned} \mathbf{H}_n^2 &= \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right) \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right) \\ &= \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T + \frac{1}{n^2} \underbrace{\mathbf{1}_n (\mathbf{1}_n^T \mathbf{1}_n) \mathbf{1}_n^T}_{=n} = \mathbf{H}_n, \end{aligned}$$

(ii) The matrix \mathbf{H}_n is a projection matrix:

$$\forall \mathbf{v} \in \mathbb{R}^n, \mathbf{H}_n \mathbf{v} = \mathbf{v} - \underbrace{\frac{1}{n} (\mathbf{v}^T \mathbf{1}_n) \mathbf{1}_n}_{= \hat{\mathbf{v}}} = \mathbf{v} - \hat{\mathbf{v}} \mathbf{1}_n,$$

where $\hat{\mathbf{v}}$ is the mean of \mathbf{v} . We can obviously observe that the matrix \mathbf{H}_n projects \mathbf{v} into the subspace of vectors that have zero mean. As we know, in linear algebra, for a vector to have zero mean implies $\mathbf{v} \perp \text{span}(\mathbf{1}_n)$. Hence, from linear algebra point of view, the matrix \mathbf{H}_n projects \mathbf{v} onto the hyperplane which is orthogonal to the all-ones vector.

B. The OLS covariance estimator: $\hat{\boldsymbol{\Sigma}}_{OLS}$

It is important to note that the estimation of the covariance matrix through optimization of an objective function (i.e. the likelihood function) can be a difficult numerical problem since the resulting estimates must be positive definite. In order to address the positivity definiteness constraint problem of $\hat{\boldsymbol{\Sigma}}_{SCM}$, it would be important to consider that the covariance matrix $\boldsymbol{\Sigma}$ is made up of $\boldsymbol{\Sigma} = \mathbf{T}^{-1} \mathbf{D} \mathbf{T}^{-T}$, where \mathbf{T} is a unit lower triangular matrix and \mathbf{D} is a diagonal matrix with positive entries [19], [17], [18].

In order to build the estimator via linear regression [17], [19], [18], let us first denote $\hat{\mathbf{x}} = [\hat{x}_1, \dots, \hat{x}_p]^T \in \mathbb{R}^p$, and consider each element \hat{x}_t , $t \in [1, p]$, as the linear least squares predictor of x_t based on its $t-1$ predecessors $\{x_j\}_{j \in [1, t-1]}$. More specifically, for $t \in [1, p]$, let

$$\hat{x}_t = \sum_{j=1}^{t-1} C_{t,j} x_j.$$

For each value of $t \geq 1$, we get

$$x_1 = \epsilon_1 ,$$

$$x_2 = C_{2,1} x_1 + \epsilon_2 ,$$

$$x_3 = C_{3,1} x_1 + C_{3,2} x_2 + \epsilon_3 ,$$

$$\vdots$$

$$x_p = C_{p,1} x_1 + C_{p,2} x_2 + \cdots + C_{p,p-1} x_{p-1} + \epsilon_p ,$$

where $\epsilon_t = x_t - \hat{x}_t$ for $t \in [1, p]$ denotes the prediction error with variance $\text{var}(\epsilon_t) = \mathbb{E}[(\epsilon_t)^2] = \theta_t^2$. Note that for $t = 1$, let $\hat{x}_1 = \mathbb{E}(x_1) = 0$, and hence, $\text{var}(\epsilon_1) = \theta_1^2 = \mathbb{E}[(x_1)^2]$.

This is equivalent to

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_p \end{pmatrix} = \begin{bmatrix} 1 & & & & \\ -C_{2,1} & 1 & & & \\ -C_{3,1} & -C_{3,2} & 1 & & \\ \vdots & \vdots & \cdots & \ddots & \\ -C_{p,1} & -C_{p,2} & \cdots & -C_{p,p-1} & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} = \mathbf{T} \mathbf{x} ,$$

where \mathbf{T} is a unit lower triangular matrix with $-C_{t,j}$ in the (t, j) th position for $t \in [2, p]$ and $j \in [1, t-1]$, and $\mathbf{x} = [x_1, x_2, \dots, x_p]^T \in \mathbb{R}^p$.

One has $\text{cov}(\boldsymbol{\epsilon}) = \mathbf{T} \boldsymbol{\Sigma} \mathbf{T}^T = \mathbf{D}$. Thus, $\mathbf{D}^{-1} = \mathbf{T}^{-T} \boldsymbol{\Sigma}^{-1} \mathbf{T}^{-1} \Rightarrow \boldsymbol{\Sigma}^{-1} = \mathbf{T}^T \mathbf{D}^{-1} \mathbf{T}$, where \mathbf{D} is a diagonal matrix with entries $\theta_1^2, \theta_2^2, \dots, \theta_p^2$. A very interesting consequence is that for any $(\hat{\mathbf{T}}, \hat{\mathbf{D}})$, $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{T}}^{-1} \hat{\mathbf{D}} \hat{\mathbf{T}}^{-T}$ is always guaranteed to be positive definite.

Let us now define the estimator $\hat{\boldsymbol{\Sigma}}_{OLS}$, where ‘‘OLS’’ stands for Ordinary Least Squares. Given a sample $\{\mathbf{x}_i\}_{i \in [1, n]}$, we have:

$$x_{i,t} = \sum_{j=1}^{t-1} C_{t,j} x_{i,j} + \epsilon_{i,t} \quad t \in [2, p], \quad i \in [1, n] .$$

Hence, for any $t \in [2, p]$, we have

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \\ \vdots \\ x_{n,t} \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,t-1} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,t-1} \\ \vdots & \vdots & & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,t-1} \end{bmatrix} \begin{bmatrix} C_{t,1} \\ C_{t,2} \\ \vdots \\ C_{t,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \vdots \\ \epsilon_{n,t} \end{bmatrix} .$$

As we can see, this is similar to a linear regression model. That is, we have

$$\mathbf{y}_t = \mathbf{A}_{n,t} \boldsymbol{\beta}_t + \mathbf{e}_t ,$$

where $\mathbf{y}_t = [x_{1,t}, \dots, x_{n,t}]^T \in \mathbb{R}^n$, $\mathbf{A}_{n,t} = [x_{i,j}]_{n \times (t-1)}$, $\boldsymbol{\beta}_t = [C_{t,1}, \dots, C_{t,t-1}]^T \in \mathbb{R}^{(t-1)}$, and $\mathbf{e}_t = [\epsilon_{1,t}, \dots, \epsilon_{n,t}]^T \in \mathbb{R}^n$.

Assume $\mathbf{e}_t \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ and $\mathbf{y}_t \sim \mathcal{N}(\mathbf{A}_{n,t} \boldsymbol{\beta}_t, \sigma^2 \mathbf{I}_n)$. Hence, by maximizing the log-likelihood under the Gaussian assumption, one can obtain the following (minimization) optimization problem:

$$\min_{\boldsymbol{\beta}_t} \left\{ \|\mathbf{y}_t - \mathbf{A}_{n,t} \boldsymbol{\beta}_t\|_2^2 \right\} . \quad (6)$$

Problem (6) has a (closed-form) solution and which is unique. This solution (optimal when $n > p$) is defined as

$$\hat{\boldsymbol{\beta}}_t^{OLS} = (\mathbf{A}_{n,t}^T \mathbf{A}_{n,t})^{-1} \mathbf{A}_{n,t}^T \mathbf{y}_t = [\hat{C}_{t,1}^{OLS}, \hat{C}_{t,t-1}^{OLS}]^T \in \mathbb{R}^{t-1} ,$$

The OLS regression coefficients $\hat{\beta}_t^{OLS}$ and the corresponding residual variance are plugged-in into \mathbf{T} and \mathbf{D} for each $t \in [2, p]$, respectively. At the end, one obtains the estimator

$$\hat{\Sigma}_{OLS} = \hat{\mathbf{T}}_{OLS}^{-1} \hat{\mathbf{D}}_{OLS} \hat{\mathbf{T}}_{OLS}^{-T}.$$

Note that $\hat{\mathbf{T}}_{OLS}$ has $-\hat{C}_{t,j}^{OLS}$ in the (t, j) th position for $t \in [2, p]$ and $j \in [1, t-1]$.

(i) $\hat{\beta}_t^{OLS}$ behaves poorly in large dimensions:

Let $\mathbf{A}_{n,t} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, with

$$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{t-1}] \in \mathbb{R}^{n \times (t-1)},$$

$$\mathbf{S} = \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_{t-1}} \end{bmatrix} \in \mathbb{R}^{(t-1) \times (t-1)},$$

and

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{t-1}] \in \mathbb{R}^{(t-1) \times (t-1)},$$

are the matrix of left singular vectors, (diagonal) matrix of singular values, and matrix of right singular vectors of $\mathbf{A}_{n,t}$, respectively. Hence, $\hat{\beta}_t^{OLS}$ can be expressed as

$$\begin{aligned} & \left((\mathbf{U}\mathbf{S}\mathbf{V}^T)^T (\mathbf{U}\mathbf{S}\mathbf{V}^T) \right)^{-1} (\mathbf{U}\mathbf{S}\mathbf{V}^T)^T \mathbf{y}_t = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^T \mathbf{y}_t \\ &= [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{t-1}] \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & & \\ & \ddots & \\ & & \frac{1}{\sqrt{\lambda_{t-1}}} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_{t-1}^T \end{bmatrix} \mathbf{y}_t \\ &= \left[\frac{\mathbf{v}_1}{\sqrt{\lambda_1}}, \frac{\mathbf{v}_2}{\sqrt{\lambda_2}}, \dots, \frac{\mathbf{v}_{t-1}}{\sqrt{\lambda_{t-1}}} \right] \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_{t-1}^T \end{bmatrix} \mathbf{y}_t \\ &= \left[\sum_{j=1}^{t-1} \frac{\mathbf{v}_j \mathbf{u}_j^T}{\sqrt{\lambda_j}} \right] \mathbf{y}_t = \sum_{j=1}^{t-1} \mathbf{z}_j, \end{aligned}$$

where $\mathbf{z}_j = \frac{\mathbf{v}_j \mathbf{u}_j^T}{\sqrt{\lambda_j}} \mathbf{y}_t$. As we can obviously observe, the singular values are located in the denominator and which causes $\hat{\beta}_t^{OLS}$ to explode in values especially when p is large enough compared to n .

By computing the mean squared error (MSE) of $\hat{\beta}_t^{OLS}$, we have

$$\begin{aligned} MSE(\hat{\beta}_t^{OLS}) &= \mathbb{E} \left[(\hat{\beta}_t^{OLS} - \beta_t)^2 \right] \\ &= \text{bias}^2(\hat{\beta}_t^{OLS}) + \text{Tr}(\text{var}(\hat{\beta}_t^{OLS})) \\ &= 0 + \sigma^2 \text{Tr}(\mathbf{A}_{n,t}^T \mathbf{A}_{n,t})^{-1} \\ &= \sigma^2 \sum_{j=1}^{t-1} \frac{1}{\lambda_j}. \end{aligned}$$

Hence, we obtain

$$\mathbb{E} \left[\left\| \hat{\beta}_t^{OLS} \right\|^2 \right] = \|\beta_t\|^2 + \sigma^2 \text{Tr}(\mathbf{A}_{n,t}^T \mathbf{A}_{n,t})^{-1}.$$

We can observe that $\|\hat{\beta}_t^{OLS}\|$ is in general larger than $\|\beta_t\|$, and thus, $\hat{\beta}_t^{OLS}$ has in general much larger coefficients (in absolute value) than the true β_t especially when p is large compared to n (or even when $p > n$).

(ii) Proving that the solution to problem (6) exists:

Let $f(\beta_t) = \|\mathbf{A}_{n,t}\beta_t - \mathbf{y}_t\|_2^2$, with $\beta_t \in Q = \mathbb{R}^{(t-1)}$.

Assume the following lemma:

Lemma II.1

Assume there $\exists (\beta_t^{(k)})_k$ and a constant $c > 0$ such that $\|\beta_t^{(k)}\|_2 \xrightarrow{k \rightarrow +\infty} +\infty$ and $\|\mathbf{A}_{n,t}\beta_t^{(k)}\|_2 \leq c \forall k$.

We have

$$\mathbf{d}_t^{(k)} = \frac{\beta_t^{(k)}}{\|\beta_t^{(k)}\|_2} \Rightarrow \|\mathbf{A}_{n,t}\mathbf{d}_t^{(k)}\|_2 = \frac{\|\mathbf{A}_{n,t}\beta_t^{(k)}\|_2}{\|\beta_t^{(k)}\|_2} \xrightarrow{k \rightarrow +\infty} 0.$$

Clearly $\mathbf{d}_t^{(k)} \in \ell(0, 1)$ is compact, and thus, we can extract a convergent sub-series towards $\mathbf{d}_t \in \ell(0, 1)$, and by continuity, we have $\|\mathbf{A}_{n,t}\mathbf{d}_t\|_2 = 0 \Rightarrow \mathbf{A}_{n,t}\mathbf{d}_t = \mathbf{0}$. As we know that $\|\mathbf{d}_t\|_2 = 1$, then Lemma II.1 does not hold, and thus, $\|\mathbf{A}_{n,t}\beta_t^{(k)}\|_2 \xrightarrow{\|\beta_t^{(k)}\|_2 \rightarrow +\infty} +\infty$.

We know that

$$\|\mathbf{A}_{n,t}\beta_t - \mathbf{y}_t\|_2^2 \geq (\|\mathbf{A}_{n,t}\beta_t\|_2 - \|\mathbf{y}_t\|_2)^2.$$

This implies $\|\mathbf{A}_{n,t}\beta_t - \mathbf{y}_t\|_2^2 \xrightarrow{\|\beta_t\|_2 \rightarrow +\infty} +\infty$, and thus, $f(\beta_t)$ is coercive. In addition $f(\beta_t)$ is continuous \Rightarrow lower semicontinuous². Hence, $f(\beta_t)$ is inf-compact. In addition, as the set Q is closed, we can confirm that there exists a solution for problem (6).

(iii) Proving that the solution that exists for problem (6) is unique:

We have $Q = \mathbb{R}^{t-1}$ is convex. Assume the following lemma:

Lemma II.2

Suppose $\gamma \in]0, 1[$, $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{t-1}$, and $\mathbf{u} \neq \mathbf{v}$.

We have

$$\begin{aligned} & \|\mathbf{A}_{n,t}(\gamma\mathbf{u} + (1-\gamma)\mathbf{v}) - \mathbf{y}_t\|_2^2 \\ &= \|\gamma(\mathbf{A}_{n,t}\mathbf{u} - \mathbf{y}_t) + (1-\gamma)(\mathbf{A}_{n,t}\mathbf{v} - \mathbf{y}_t)\|_2^2 \\ &\leq (\gamma\|\mathbf{A}_{n,t}\mathbf{u} - \mathbf{y}_t\|_2 + (1-\gamma)\|\mathbf{A}_{n,t}\mathbf{v} - \mathbf{y}_t\|_2)^2 \\ &\leq \gamma\|\mathbf{A}_{n,t}\mathbf{u} - \mathbf{y}_t\|_2^2 + (1-\gamma)\|\mathbf{A}_{n,t}\mathbf{v} - \mathbf{y}_t\|_2^2. \end{aligned}$$

Hence, $f(\beta_t)$ is convex. In order for the equality to hold, we should have the following:

- $\gamma(\mathbf{A}_{n,t}\mathbf{u} - \mathbf{y}_t) = k(1-\gamma)(\mathbf{A}_{n,t}\mathbf{v} - \mathbf{y}_t)$, with $k \geq 0$;
- $\|\mathbf{A}_{n,t}\mathbf{u} - \mathbf{y}_t\|_2 = \|\mathbf{A}_{n,t}\mathbf{v} - \mathbf{y}_t\|_2$, and thus, $\frac{k(1-\gamma)}{\gamma} = 1$.

Assume $\mathbf{A}_{n,t}\mathbf{u} = \mathbf{A}_{n,t}\mathbf{v} \Rightarrow \mathbf{u} = \mathbf{v}$. Hence, Lemma II.2 does not hold and $f(\beta_t)$ is strictly convex. We can confirm that the solution that exists for problem (6) is unique.

²For a continuous function f and $S_\eta = \{\beta_t \in Q / f(\beta_t) \leq \eta\} = f^{-1}([-\infty, \eta])$, $\forall \eta \in \mathbb{R}$. The inverse image of a closed set by a continuous application is closed, and thus, S_η is closed $\Rightarrow f$ is lower semicontinuous.

C. The Eigenvalue Clipping covariance estimator: $\hat{\Sigma}_{Clipping}$

The proposed idea by Leplou et al. [21] was that all the eigenvalues beyond $(1 + \sqrt{\frac{p}{n}})^2$ are interpreted as signal while the others as a pure noise [22], [23]. More precisely, the Sample covariance matrix, $\hat{\Sigma}_{SCM}$, is standardized and all its eigenvectors are kept unchanged. Denote the standardized version by the matrix $\tilde{\Sigma}_{SCM}$. Let $\tilde{\Sigma}_{SCM} = \tilde{\mathbf{Q}}\tilde{\Lambda}\tilde{\mathbf{Q}}^T$, with

$$\tilde{\mathbf{Q}} = [\tilde{\mathbf{q}}_1, \dots, \tilde{\mathbf{q}}_p] \in \mathbb{R}^{p \times p}$$

and

$$\tilde{\Lambda} = \begin{bmatrix} \tilde{\lambda}_1 & & \\ & \ddots & \\ & & \tilde{\lambda}_p \end{bmatrix} \in \mathbb{R}^{p \times p}$$

are the matrix of eigenvectors and diagonal matrix of eigenvalues of $\tilde{\Sigma}_{SCM}$, respectively.

The eigenvalue clipping covariance matrix estimator, $\hat{\Sigma}_{Clipping}$, can be computed as

$$\hat{\Sigma}_{Clipping} = \sum_{j=1}^p \lambda_j^* \tilde{\mathbf{q}}_j \tilde{\mathbf{q}}_j^T,$$

with

$$\lambda_j^* = \begin{cases} \tilde{\lambda}_j & \text{if } \tilde{\lambda}_j > (1 + \sqrt{\frac{p}{n}})^2 \\ \bar{\lambda} & \text{otherwise} \end{cases},$$

where $\bar{\lambda}$ has to be chosen such as $\text{Tr}(\tilde{\Sigma}_{SCM}) = \text{Tr}(\hat{\Sigma}_{Clipping})$.

III. EXPERIMENTS

A. Our procedure for the selection of assets from the Russell1000 dataset

We select a window time from 2006 onward: 85% of stocks with full data values. In order to build a K -assets portfolio, on the training set, we perform the following:

- Clustering based on asymmetry with K clusters;
- In each cluster, we select the asset with the best Sharpe ratio.

The main advantages of such a procedure is that it provides (1) a large pool of diversified stocks with interesting historical Sharpe ratios; (2) a stability gain for the covariance matrix estimators.

For any kind of portfolio optimization problem, we normally compute the optimal parameters for a given sample data (i.e., the portfolio returns for the last 2000 days). In order to run our strategy (with a particular covariance matrix estimator) over a given data, we do it in-sample where we expect to “technically” obtain the best possible result. However, by evaluating our covariance estimation strategy over a different period that was used for the in-sample data (i.e., the next 252 days), then we have to do it out-of-sample. That is, we are evaluating the covariance estimation strategy on a period which is different from the one we optimized the strategy on. Our evaluation (as a back-testing) on the out-of-sample will be quite important mainly because we will only be able to invest today the result of the optimization that has been made today. Hence, it is like we are earning tomorrow’s strategy performance and not the one of yesterday.

B. Calculating probabilities between $\hat{\Sigma}_{SCM}$ and $\hat{\Sigma}_{OLS}$

In order to compare between the fraction of times [24] that the risk of $\hat{\Sigma}_{SCM}$ is smaller than $\hat{\Sigma}_{OLS}$ and vice-versa, it would be useful to consider the realized (out-of-sample) risk as our based performance measure. We define the realized risk [24] as

$$R_V^{out} = \sqrt{252 * \hat{\mathbf{w}}^T \Sigma^{out} \hat{\mathbf{w}}},$$

where Σ^{out} is the out-of-sample covariance matrix, and $\hat{\mathbf{w}}$ are computed from the in-sample covariance matrix (filtered or not).

By “fraction of time”, we mean that we need to calculate the probability that $\hat{\Sigma}_{SCM}$ outperforms $\hat{\Sigma}_{OLS}$ and vice-versa for a particular in-sample window size (i.e. $n = 30, n = 50, n = 100, n = 200, n = 500$, or $n = 800$). To do that, we will perform 20 independent simulations of $p = 100$ assets in random periods.

From Figure 1, we can observe that both $\hat{\Sigma}_{SCM}$ and $\hat{\Sigma}_{OLS}$ perform roughly the same but $\hat{\Sigma}_{SCM}$ is in general slightly better than $\hat{\Sigma}_{OLS}$ for our specific dataset assets’ order.

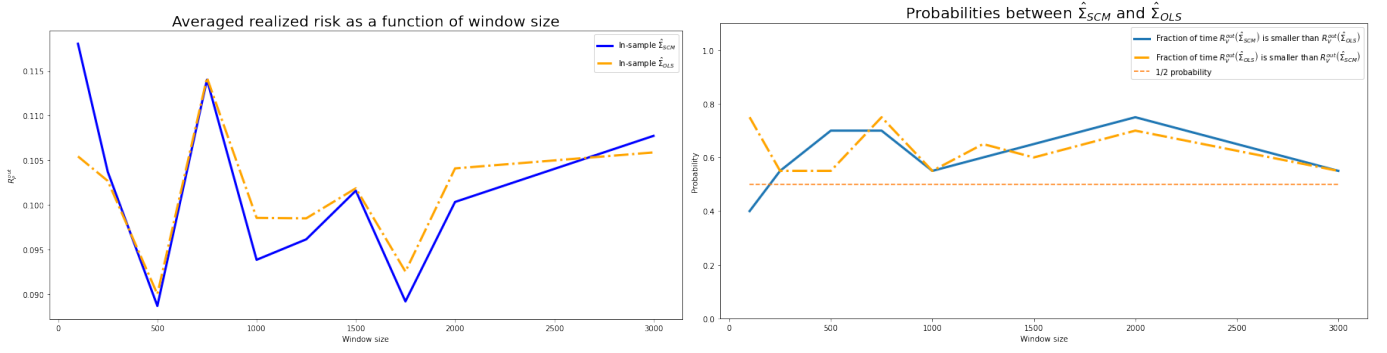


Fig. 1. Probabilities between $\hat{\Sigma}_{SCM}$ and $\hat{\Sigma}_{OLS}$.

C. Comparison between $\hat{\Sigma}_{SCM}$, $\hat{\Sigma}_{OLS}$, $\hat{\Sigma}_{CV}$, and $\hat{\Sigma}_{Clipping}$

1) *Eigenvalue stability*: In order to compare the best (unknown) covariance estimator (aka oracle estimator) with $\hat{\Sigma}_{SCM}$, $\hat{\Sigma}_{OLS}$, $\hat{\Sigma}_{CV}$, and $\hat{\Sigma}_{Clipping}$, it would be important to compare the in- and out-of-sample persistence of the eigenvalues of these covariance estimators [24], [25].

As the analytical solution of the optimal portfolio weight calculated via the Markowitz optimization problem (that is, the Markowitz's formulation in terms of a soft constraint with regards to aversion to risk) mainly depends on both the (unknown) covariance matrix Σ and its inverse Σ^{-1} , thus we are going to consider two measures in order to highlight on both the small and large eigenvalues [24], [25].

High Eigenvalues Stability measure:

$$\epsilon_{HighEigen} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\lambda_i - z_i)^2}$$

Low Eigenvalues Stability measure:

$$\epsilon_{LowEigen} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{1}{\lambda_i} - \frac{1}{z_i} \right)^2}$$

where we have λ_i defined as the i^{th} eigenvalue of the in-sample (filtered or unfiltered) covariance estimator, and z_i is the i^{th} eigenvalue for the oracle estimator. As we can observe, the measure $\epsilon_{HighEigen}$ must account for the discrepancy between the highest eigenvalues, whereas the measure $\epsilon_{LowEigen}$ has to attribute more weight to the discrepancy between the smallest eigenvalues [24].

The main idea is to plot those measures as a function of the window size and compare which covariance estimator is closer to the oracle estimator. As can be seen from Figure 2, $\hat{\Sigma}_{Clipping}$ is the closest to the oracle estimator for the high eigenvalues stability measure.

2) *Some metrics*: we will use the following metrics: (1) expected risk; (2) gross leverage; (3) log Sharpe ratio; and (4) effective diversification. From Figure 3, we can obviously see the superiority of $\hat{\Sigma}_{Clipping}$ with respect to $\hat{\Sigma}_{SCM}$, $\hat{\Sigma}_{OLS}$, and $\hat{\Sigma}_{CV}$ for all the four metrics.

D. Penalizing turnovers

We define the turnover as

$$\frac{1}{\Delta_\tau} \sum_{\tau=1}^{\Delta_\tau-1} \sum_{j=1}^p |w_j^{(\tau)} - w_j^{(\tau+1)}|.$$

It is known that the effective portfolio diversification can be defined as the ratio $\frac{1}{\sum_{j=1}^p w_j^2}$ that represents the effective number of stocks with a significant amount of money invested. However, this ratio does not have a clear interpretation when short-selling is allowed. In this regard, we may assume the following alternative metric for the effective diversification

$$N_q = \underset{l}{\operatorname{argmin}} \sum_{i=1}^l |w_i| \geq q \sum_{j=1}^p |w_j|,$$

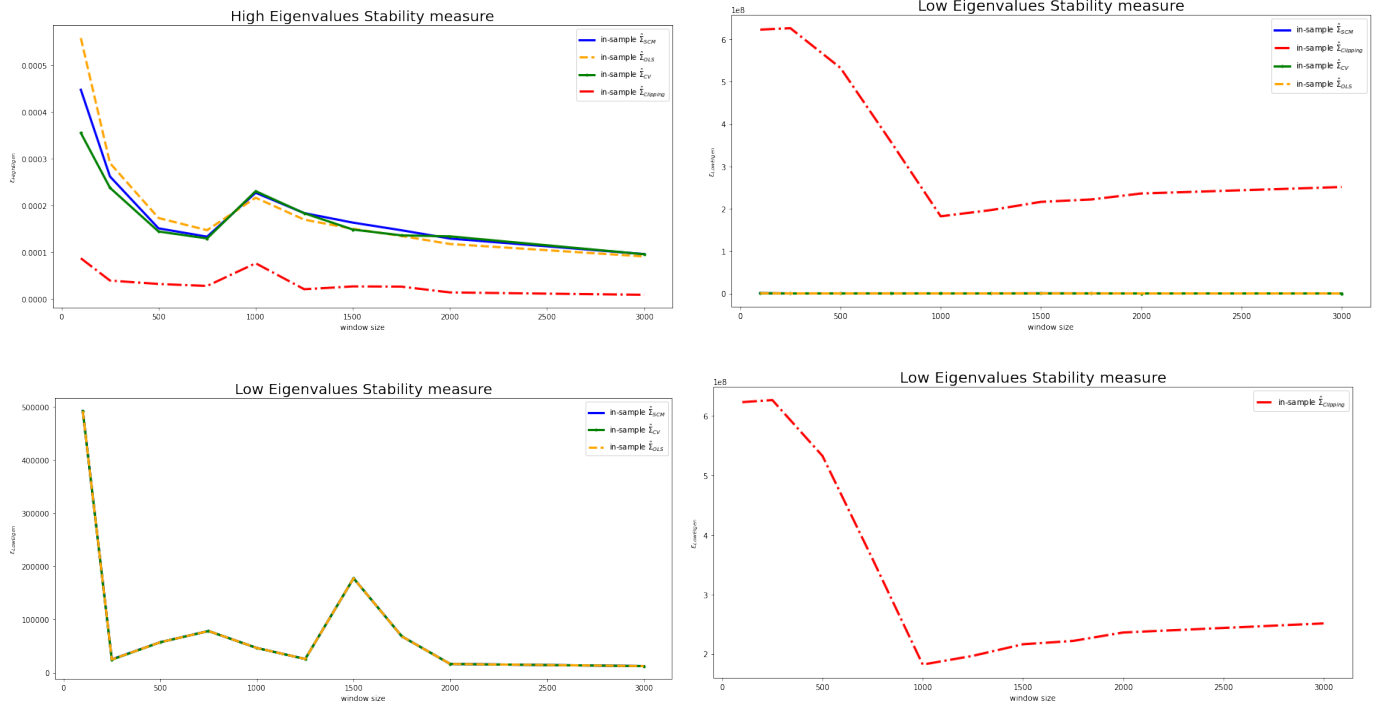


Fig. 2. High vs Low eigenvalue stability.

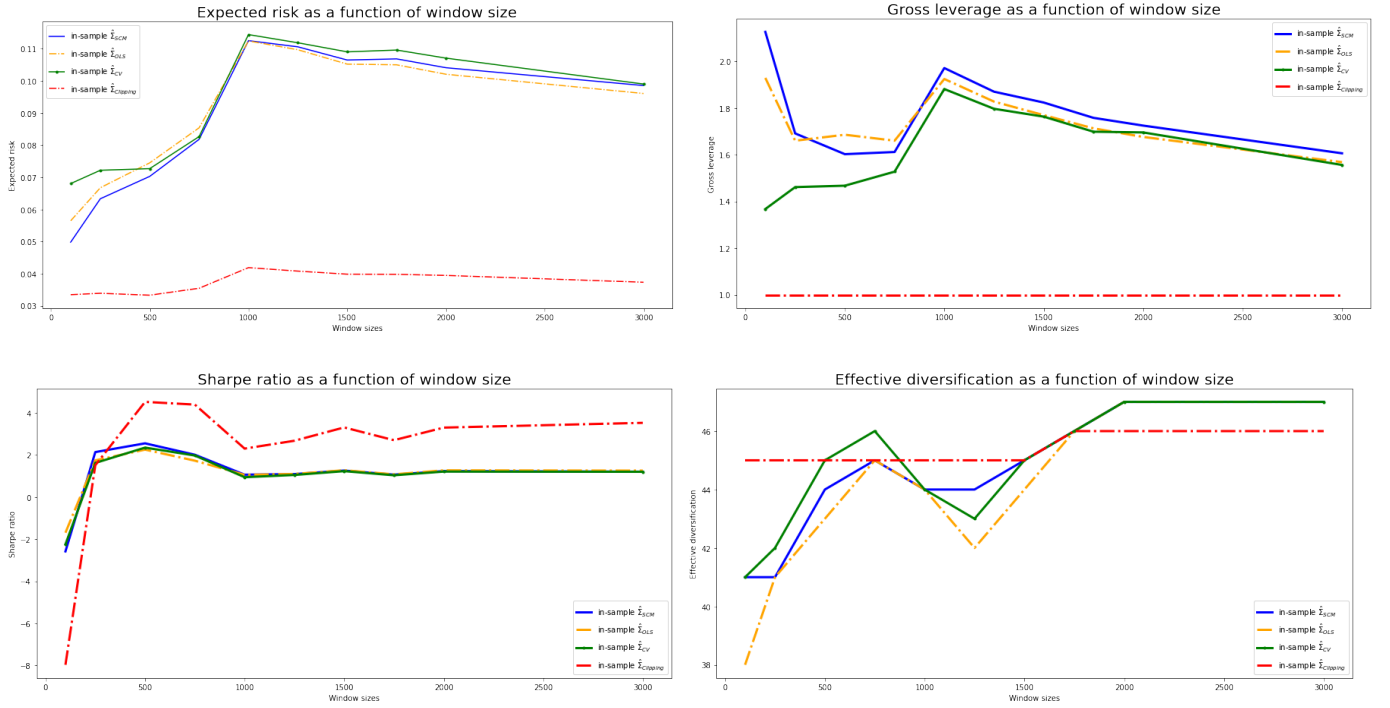


Fig. 3. From top left to bottom right: expected risk, gross leverage, log Sharpe ratio, and effective diversification.

which is the number of stocks that accounts for ‘ q ’ percent of the total amount of money invested (i.e., $q = 90\%$ may be a good choice). As we can observe from Figure 4, the l_1 -norm is more stringent than the l_2 -norm but converges much slower.

E. Portfolio Investment Policy

We define two strategies:

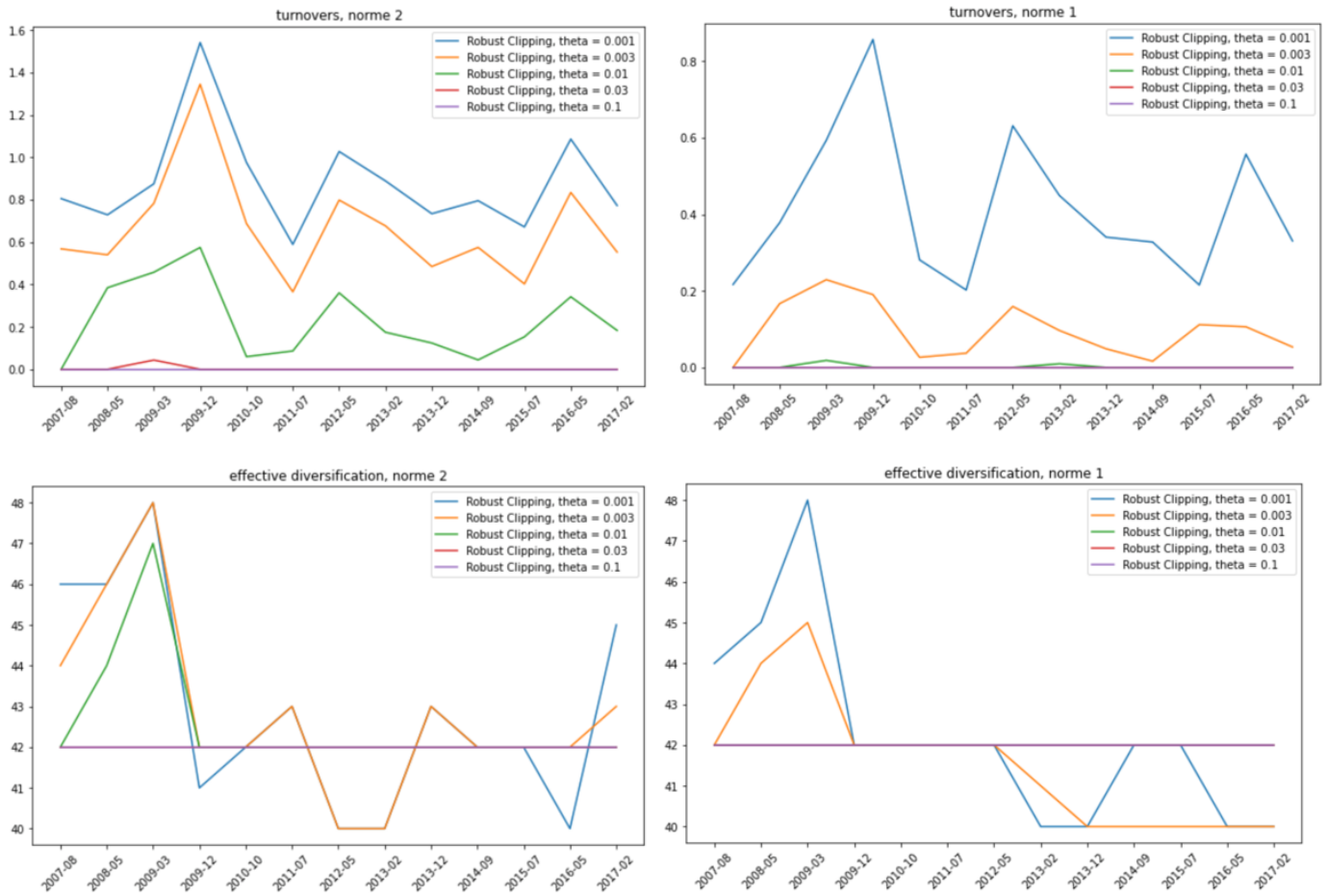


Fig. 4. Results obtained with robust portfolio.

- 1 very reactive strategy (client B):
 - In-sample window size = 2 months;
 - Out-sample window size = 1 month;
 - AIM: capture short-term variations of the market.
- 1 long-term strategy (client A):
 - In-sample window size = 1 year;
 - Out-sample window size = 6 months;
 - AIM: Long-term investment portfolios, less rebalancing costs.
- Trade-offs expressed in soft constraints;
- Rebalancing costs will be penalized.

For client A: We assume the following needs: (1) a risk-averse; (2) an averse to rebalancing fees; and (3) invest with sparsity.

1) Robust portfolio VS minimum-variance with eigenvalue clipping (with $n > p$):

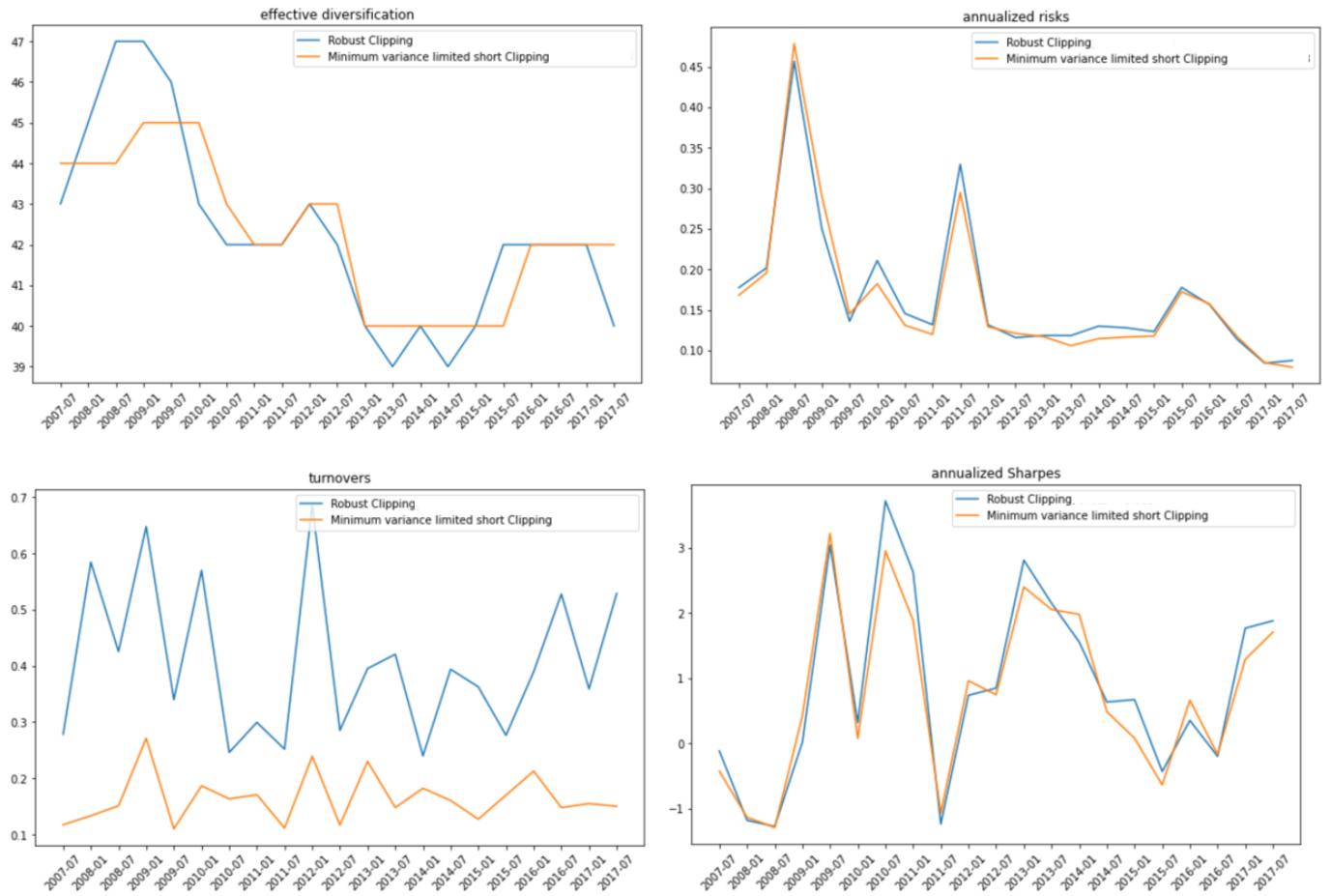


Fig. 5. Robust portfolio VS minimum-Variance with eigenvalue clipping.

- Parameters:
 - 50 diversified stocks;
 - In-Sample = 252 days;
 - Out-Sample = 126 days;
 - $\theta = 0.003$;
 - $\beta = 0.00001$;
 - $\lambda = 0.5$;
 - $\kappa = 0.5$;
 - $\min_short = -0.5$;
 - $\Omega = \sqrt{\Sigma}$.

As we can observe from Figure 5, with similar risks, the minimum-variance portfolio has less rebalancing fees than the robust portfolio.

2) Equi-risk budgeting with eigenvalue clipping (with $n > p$):

- Parameters:
 - 10 diversified stocks;
 - In-Sample = 252 days;
 - Out-Sample = 126 days.

From Figure 6, we can observe that the equi-risk budgeting with eigenvalue clipping, unless specified by the client, is a less advantageous strategy, less diversified portfolio with higher risks, and less robust than the robust portfolio with eigenvalue clipping.

For client B: We assume the following needs: (1) risk-taking; (2) willing to pay rebalancing fees; and (3) follow short-term market's movements.

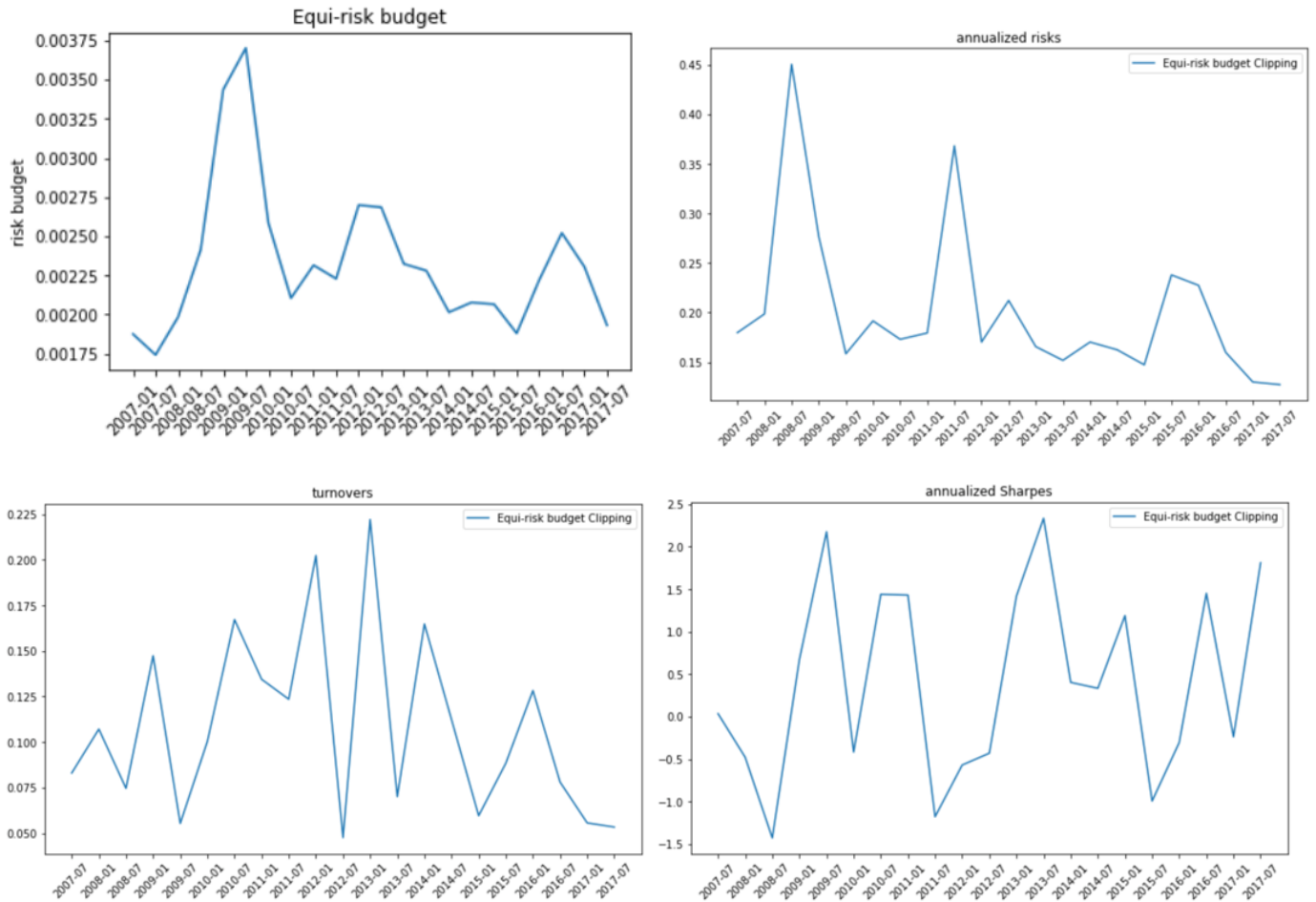


Fig. 6. Equi-risk budgeting with eigenvalue clipping.

1) Robust portfolio VS mean-variance with eigenvalue clipping (with $n > p$):

- Parameters:
 - 50 diversified stocks;
 - In-Sample = 2 months;
 - Out-Sample = 1 month;
 - $\theta = 0.0005$;
 - $\alpha = 0.0005$;
 - $\lambda = 0.01$;
 - $\kappa = 0.5$;
 - $\Omega = \sqrt{\Sigma}$.

From Figure 7, we can conclude that the mean-variance is not a good model to follow. On the other hand, the robust portfolio keeps a highly diversified portfolio as well as it achieves higher turnovers and Sharpe ratios (in average) than the risk-based portfolios (minimum-variance and equi-risk budgeting).

From all the results above, we can conclude that robust portfolio with eigenvalue clipping is the model suitable for client B, whereas minimum-variance with eigenvalue clipping is the one suitable for client A.

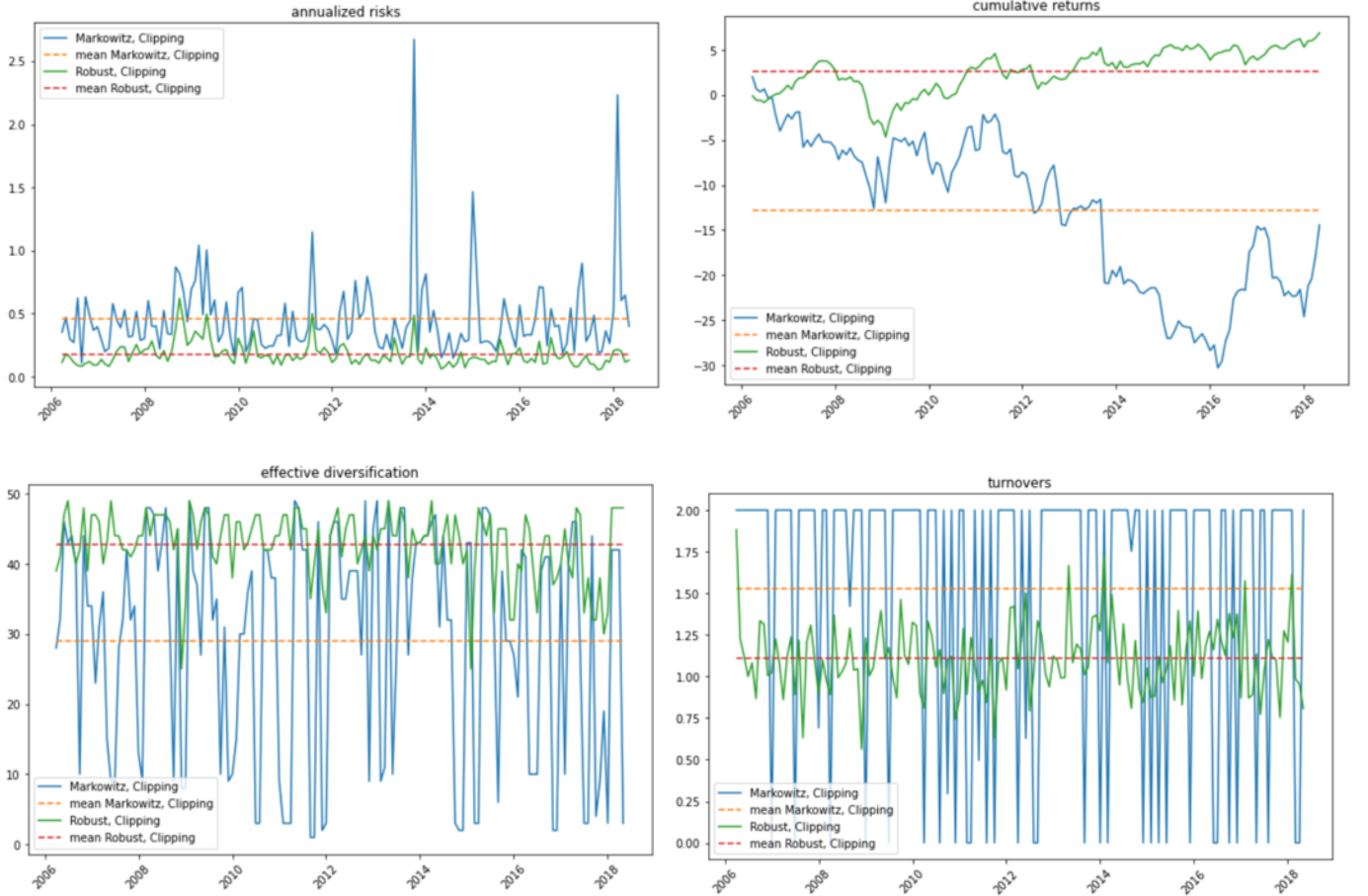


Fig. 7. Robust portfolio VS mean-variance with eigenvalue clipping.

IV. CONCLUSION AND FUTURE WORK

A. Conclusion

In this technical report, we combined four different portfolio optimization strategies (mean-variance, robust portfolio, minimum-variance, and equi-risk budgeting) with four covariance matrix estimators (sample covariance matrix, OLS covariance matrix, cross-validated eigenvalue shrinkage covariance matrix, and eigenvalue clipping) to meet two types of clients' requirements: client A who wants to invest money wisely, not taking too much risk, and not willing to pay too much in rebalancing fees; and client B who wants to make money quickly, benefit from market's short-term volatility, and ready to pay rebalancing fees. Our experiments on the Russell1000 dataset suggest that the robust portfolio with eigenvalue clipping is the strategy suitable for client B, whereas minimum-variance with eigenvalue clipping is the one suitable for client A.

B. Some directions for future work

As for future enhancements, a likely first step would be to extend this work by combining the four portfolio optimization strategies with covariance matrix estimators in the case when $p > n$. For example, and as was mentioned in section II-B, $\hat{\beta}_t^{OLS}$ behaves poorly in large dimensions, and thus, one can simply think to penalize it (e.g. by the l_2 norm). In addition, it would be important to propose robust (i.e. distributional-free estimators) and/or sparse covariance estimators and combine them with the robust portfolio optimization discussed in problem (3).

Penalizing the least squares estimator $\hat{\beta}_t^{OLS}$: Recall that we have $\hat{\beta}_t^{OLS} = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^T\mathbf{y}_t = \sum_{j=1}^{t-1} \mathbf{z}_j$, where $\mathbf{z}_j = \frac{\mathbf{v}_j \mathbf{u}_j^T}{\sqrt{\lambda_j}} \mathbf{y}_t$.

A general penalization of $\hat{\beta}_t^{OLS}$ can be written in the form

$$\hat{\beta}_t^{\text{pen}} = \mathbf{V}\mathbf{S}^{-1}\mathbf{\Gamma}\mathbf{U}^T\mathbf{y}_t = \sum_{j=1}^{t-1} g(\lambda_j) \mathbf{z}_j,$$

where

$$\mathbf{\Gamma} = \begin{bmatrix} g(\lambda_1) & & \\ & \ddots & \\ & & g(\lambda_{t-1}) \end{bmatrix} \in \mathbb{R}^{(t-1) \times (t-1)},$$

and $g(\lambda_j)$ is the penalization factor for $j \in [1, t-1]$.

As we can clearly observe, $\hat{\beta}_t^{\text{pen}}$ reduces to $\hat{\beta}_t^{\text{OLS}}$ when $g(\lambda_j) = 1, \forall j \in [1, t-1]$.

Some properties about $\hat{\beta}_t^{\text{pen}}$:

We have

- $\mathbb{E}(\hat{\beta}_t^{\text{pen}}) = \mathbf{VS}^{-1}\mathbf{\Gamma U}^T \mathbb{E}(\mathbf{y}_t) = \mathbf{VS}^{-1}\mathbf{\Gamma S V}^T \beta_t.$

This implies

$$\begin{aligned} \text{bias}^2(\hat{\beta}_t^{\text{pen}}) &= \left(\mathbb{E}(\hat{\beta}_t^{\text{pen}} - \beta_t) \right)^T \mathbb{E}(\hat{\beta}_t^{\text{pen}} - \beta_t) \\ &= \left(\mathbf{VS}^{-1}\mathbf{\Gamma S V}^T \beta_t - \beta_t \right)^T \left(\mathbf{VS}^{-1}\mathbf{\Gamma S V}^T \beta_t - \beta_t \right) \\ &= \left(\mathbf{VS}^{-1}\mathbf{\Gamma S V}^T \beta_t - \mathbf{V V}^T \beta_t \right)^T \left(\mathbf{VS}^{-1}\mathbf{\Gamma S V}^T \beta_t - \mathbf{V V}^T \beta_t \right) \\ &= \left(\mathbf{V}(\mathbf{S}^{-1}\mathbf{\Gamma S} - \mathbf{I}) \mathbf{V}^T \beta_t \right)^T \left(\mathbf{V}(\mathbf{S}^{-1}\mathbf{\Gamma S} - \mathbf{I}) \mathbf{V}^T \beta_t \right) \\ &= \left(\mathbf{V}^T \beta_t \right)^T (\mathbf{S}^{-1}\mathbf{\Gamma S} - \mathbf{I})^T (\mathbf{S}^{-1}\mathbf{\Gamma S} - \mathbf{I}) \left(\mathbf{V}^T \beta_t \right) \\ &= \sum_{j=1}^{t-1} (g(\lambda_j) - 1)^2 \left(\mathbf{v}_j^T \beta_t \right)^2. \end{aligned}$$

- $\text{var}(\hat{\beta}_t^{\text{pen}}) = (\mathbf{VS}^{-1}\mathbf{\Gamma U}^T) (\mathbf{VS}^{-1}\mathbf{\Gamma U}^T)^T \text{var}(\mathbf{y}_t) = \sigma^2 \mathbf{VS}^{-1}\mathbf{\Gamma U}^T \mathbf{U} \mathbf{\Gamma S}^{-1} \mathbf{V}^T.$

This implies

$$\text{Tr}(\text{var}(\hat{\beta}_t^{\text{pen}})) = \sigma^2 \text{Tr}(\mathbf{VS}^{-1}\mathbf{\Gamma}^2 \mathbf{S}^{-1} \mathbf{V}^T) = \sigma^2 \sum_{j=1}^{t-1} \frac{(g(\lambda_j))^2}{\lambda_j}.$$

By computing the mean squared error (MSE) of $\hat{\beta}_t^{\text{pen}}$, we have

$$\text{MSE}(\hat{\beta}_t^{\text{pen}}) = \mathbb{E} \left[\left(\hat{\beta}_t^{\text{pen}} - \beta_t \right)^2 \right] = \sum_{j=1}^{t-1} (g(\lambda_j) - 1)^2 \left(\mathbf{v}_j^T \beta_t \right)^2 + \sigma^2 \sum_{j=1}^{t-1} \frac{(g(\lambda_j))^2}{\lambda_j}.$$

We can clearly observe that one needs to set $g(\lambda_j) \neq 1$ and more precisely $|g(\lambda_j)| < 1$ in order to have lower variance than $\hat{\beta}_t^{\text{OLS}}$.

ACKNOWLEDGMENT

We would like to thank Christian Bonjorno (CentraleSupélec) and Romain Perchet (BNP Paribas Asset Management) for their time in teaching us the course Portfolio allocation at CentraleSupélec in 2021.

REFERENCES

- 1 Sufi, A. and Taylor, A. M., "Financial crises: A survey," National Bureau of Economic Research, Working Paper 29155, August 2021. [Online]. Available: <http://www.nber.org/papers/w29155>
- 2 Markowitz, H., "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952. [Online]. Available: <http://www.jstor.org/stable/2975974>
- 3 Perchet, R., Xiao, L., de Carvalho, R. L., and Heckel, T., "Insights into robust portfolio optimization: Decomposing robust portfolios into mean-variance and risk-based portfolios," *Econometric Modeling: Capital Markets - Portfolio Theory eJournal*, 2015.
- 4 Chopra, V. K., "Improving optimization," *The Journal of Investing*, vol. 2, no. 3, pp. 51–59, 1993. [Online]. Available: <https://joi.pm-research.com/content/2/3/51>
- 5 Frost, P. A. and Savarino, J. E., "For better performance," *The Journal of Portfolio Management*, vol. 15, no. 1, pp. 29–34, 1988. [Online]. Available: <https://jpm.pm-research.com/content/15/1/29>
- 6 Chopra, V. K. and Ziemba, W. T., "The effect of errors in means, variances, and covariances on optimal portfolio choice," *The Journal of Portfolio Management*, vol. 19, no. 2, pp. 6–11, 1993. [Online]. Available: <https://jpm.pm-research.com/content/19/2/6>
- 7 Black, F. and Litterman, R. B., "Asset allocation," *The Journal of Fixed Income*, vol. 1, no. 2, pp. 7–18, 1991. [Online]. Available: <https://jfi.pm-research.com/content/1/2/7>
- 8 Yin, C., Perchet, R., and Soupé, F., "A practical guide to robust portfolio optimization," *Quantitative Finance*, vol. 21, no. 6, pp. 911–928, 2021. [Online]. Available: <https://doi.org/10.1080/14697688.2020.1849780>

- 9 Lobo, M. S., Vandenberghe, L., Boyd, S., and Lebret, H., "Applications of second-order cone programming," *Linear Algebra and its Applications*, vol. 284, no. 1, pp. 193–228, 1998, international Linear Algebra Society (ILAS) Symposium on Fast Algorithms for Control, Signals and Image Processing. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0024379598100320>
- 10 Tütüncü, R. H. and Koenig, M., "Robust asset allocation," *Annals of Operations Research*, vol. 132, pp. 157–187, 2004.
- 11 Ceria, S. and Stubbs, R. A., *Incorporating Estimation Errors into Portfolio Selection: Robust Portfolio Construction*. Cham: Springer International Publishing, 2016, pp. 270–294. [Online]. Available: https://doi.org/10.1007/978-3-319-30794-7_12
- 12 B., S., "Can robust portfolio optimization help to build better portfolios?" *Journal of Asset Management*, vol. 7, pp. 374–387, 2006.
- 13 de Carvalho, R. L., Lu, X., and Moulin, P., "Demystifying equity risk-based strategies: A simple alpha plus beta description," *The Journal of Portfolio Management*, vol. 38, pp. 56 – 70, 2012.
- 14 Maillard, S., Roncalli, T., and Teïletche, J., "The properties of equally weighted risk contribution portfolios," *The Journal of Portfolio Management*, vol. 36, no. 4, pp. 60–70, 2010. [Online]. Available: <https://jpm.pm-research.com/content/36/4/60>
- 15 Clarke, R., Silva, H. d., and Thorley, S., "Risk parity, maximum diversification, and minimum variance: An analytic perspective," *The Journal of Portfolio Management*, vol. 39, no. 3, pp. 39–53, 2013. [Online]. Available: <https://jpm.pm-research.com/content/39/3/39>
- 16 Kempf, A. and Memmel, C., "On the estimation of the global minimum variance portfolio," Tech. Rep., 2005.
- 17 Bitar, A. W., Ovarlez, J.-P., and Cheong, L.-F., "Sparsity-based cholesky factorization and its application to hyperspectral anomaly detection," in *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2017, pp. 1–5.
- 18 Bitar, A. W., Ovarlez, J. P., and Cheong, L. F., "Exploitation de la parcimonie par la factorisation de cholesky et son application pour la détection d'anomalies en imagerie hyperspectrale," 2017.
- 19 Pourahmadi, M., "Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation," *Biometrika*, vol. 86, no. 3, pp. 677–690, 1999. [Online]. Available: <http://www.jstor.org/stable/2673662>
- 20 Bartz, D., "Cross-validation based nonlinear shrinkage," 2016. [Online]. Available: <https://arxiv.org/abs/1611.00798>
- 21 Laloux, L., Cizeau, P., Potters, M., and Bouchaud, J.-P., "Random matrix theory and financial correlations," *International Journal of Theoretical and Applied Finance (IJTAF)*, vol. 03, no. 03, pp. 391–397, 2000. [Online]. Available: <https://EconPapers.repec.org/RePEc:wsj:ijtafx:v:03:y:2000:i:03:n:s0219024900000255>
- 22 Bun, J., Bouchaud, J.-P., and Potters, M., "Cleaning large correlation matrices: Tools from random matrix theory," *Physics Reports*, vol. 666, pp. 1–109, jan 2017. [Online]. Available: <https://doi.org/10.1016%2Fj.physrep.2016.10.005>
- 23 Bouchaud, J.-P. and Potters, M., "824Financial applications of random matrix theory: a short review," in *The Oxford Handbook of Random Matrix Theory*. Oxford University Press, 09 2015. [Online]. Available: <https://doi.org/10.1093/oxfordhb/9780198744191.013.40>
- 24 Bongiorno, C. and Challet, D., "Covariance matrix filtering with bootstrapped hierarchies," *PLoS ONE*, vol. 16, 2020.
- 25 Bongiorno, C., Challet, D., and Loeper, G., "Cleaning the covariance matrix of strongly nonstationary systems with time-independent eigenvalues," 2021. [Online]. Available: <https://arxiv.org/abs/2111.13109>