Regularization of Portfolio Allocation

Article *in* SSRN Electronic Journal · May 2013

DOI: 10.2139/ssrn.2767358

4 authors:

Benjamin Bruder
Paris Diderot University

Nicolas Gaussel
Metori Capital Management

J.-C. Richard
Finance

Thierry Roncalli
Université Paris-Saclay

Regularization of Portfolio Allocation*

Benjamin Bruder
Quantitative Research
Lyxor Asset Management, Paris
benjamin.bruder@lyxor.com

Nicolas Gaussel
Chief Investment Officer
Lyxor Asset Management, Paris
nicolas.gaussel@lyxor.com

Jean-Charles Richard
Quantitative Research
Lyxor Asset Management, Paris
jean-charles.richard@lyxor.com

Thierry Roncalli
Quantitative Research
Lyxor Asset Management, Paris
thierry.roncalli@lyxor.com

June 2013

Abstract

The mean-variance optimization (MVO) theory of Markowitz (1952) for portfolio selection is one of the most important methods used in quantitative finance. This portfolio allocation needs two input parameters, the vector of expected returns and the covariance matrix of asset returns. This process leads to estimation errors, which may have a large impact on portfolio weights. In this paper we review different methods which aim to stabilize the mean-variance allocation. In particular, we consider recent results from machine learning theory to obtain more robust allocation.

Keywords: Portfolio optimization, active management, estimation error, shrinkage estimator, resampling methods, eigendecomposition, norm constraints, Lasso regression, ridge regression, information matrix, hedging portfolio, sparsity.

JEL classification: G11, C60.

1 Introduction

The mean-variance optimization (MVO) framework developed by Markowitz (1952) is certainly the most famous model used in asset management. This model is generally associated to the CAPM theory of Sharpe (1964). This explains why Harry M. Markowitz and William F. Sharpe have shared[1] the Nobel Prize in 1990. However, the two models are used differently by practitioners.

The CAPM theory considers the Markowitz model from the viewpoint of micro analysis in order to deduce the price formation for financial assets. In this model, the key concept is the market portfolio, which is uniquely defined. In the Markowitz model, optimized portfolios depend on expected returns and risks. Moreover, the optimal portfolio is not unique and depends on the investor's risk aversion. As a consequence, these two models

*We are grateful to Clément Le Bars for his helpful comments.
[1] with Merton H. Miller.

pursue different purposes. While the CAPM theory is the foundation framework of passive management, the Markowitz model is the relevant framework for active management.

Nevertheless, even if the Markowitz model is a powerful model to transform the views of the portfolio manager into investment bets, it has suffered a lot of criticism because it is particularly dependent on estimation errors (Michaud, 1989). In fact, the Markowitz model is an aggressive model of active management (Roncalli, 2013). By construction, it does not make the distinction between real arbitrage factors and noisy arbitrage factors. The goal of portfolio regularization is then to produce less aggressive portfolios by reducing noisy bets.

The paper is organized as follows. Section two presents the motivations to use regularization methods. In particular, we illustrate the instability of mean-variance optimized portfolios. In section three, we review the different approaches of portfolio regularization. They concern the introduction of weight constraints, the use of resampling techniques or the shrinkage of covariance matrices. We also consider penalization methods of the objective function like Lasso or ridge regression and show how these methods may be used to regularize the inverse of the covariance matrix, which is the most important quantity in portfolio optimization. In section four, we consider different applications in order to illustrate the impact of regularization on portfolio optimization. Section five offers some concluding remarks.

2 Motivations

2.1 The mean-variance portfolio

Let us consider a universe of n risky assets. Let μ and Σ be the vector of expected returns and the covariance matrix of asset returns[2]. We note r the risk-free asset. A portfolio allocation consists in a vector of weights $x = (x_1, \ldots, x_n)$ where x_i is the percentage of the wealth invested in the i^{th} asset. Sometimes, we may assume that all the wealth is invested meaning that the sum of weights is equal to one. Moreover, we may also add some other constraints on the weights. For instance, we may impose that the portfolio is long-only. Let us now define the quadratic utility function $\mathcal{U}$ of the investor which only depends of the expected returns μ and the covariance matrix Σ of the assets:

$$\mathcal{U}(x) = x^\top (\mu - r\mathbf{1}) - \frac{\phi}{2} x^\top \Sigma x$$

where ϕ is the risk tolerance of the investor. The mean-variance optimized (or MVO) portfolio $x^\star$ is the portfolio which maximizes the investor's utility. The optimization problem can be reformulated equivalently as a standard QP problem:

$$x^\star = \arg\min \frac{1}{2} x^\top \Sigma x - \gamma x^\top (\mu - r\mathbf{1})$$

where $\gamma = \phi^{-1}$. Without any constraints, the solution yields the well known formula:

$$x^\star = \frac{1}{\phi} \Sigma^{-1} (\mu - r\mathbf{1}) = \gamma \Sigma^{-1} (\mu - r\mathbf{1})$$

It comes that the Sharpe ratio of the MVO portfolio is:

$$\begin{aligned} \mathrm{SR}(x^\star \mid r) &= \frac{x^{\star\top} (\mu - r\mathbf{1})}{\sqrt{x^{\star\top} \Sigma x^\star}} \\ &= \sqrt{(\mu - r\mathbf{1})^\top \Sigma^{-1} (\mu - r\mathbf{1})} \end{aligned}$$

[2]In this paper, we adopt the formulation presented in the book of Roncalli (2013).

We deduce that the optimal utility of the investor is:

$$\begin{aligned}
\mathcal{U}(x^\star) &= x^{\star\top}(\mu - r\mathbf{1}) - \frac{\phi}{2} x^{\star\top} \Sigma x^\star \\
&= \frac{1}{2\phi} (\mu - r\mathbf{1})^\top \Sigma^{-1} (\mu - r\mathbf{1}) \\
&= \frac{1}{2\phi} \mathrm{SR}^2 (x^\star \mid r)
\end{aligned}$$

Maximizing the mean-variance utility function is then equivalent to maximizing the ex-ante Sharpe ratio of the allocation.

Remark 1 *Without lack of generality, we assume that the risk-free rate r is equal to zero in the rest of the paper.*

In practice, we cannot reach the optimal allocation because we don't know μ and Σ. That is why we have to estimate these two quantities. Let $R_t = (R_{1,t}, \ldots, R_{n,t})$ be the vector of historical returns for the different assets at time t. We then estimate μ and Σ by maximum likelihood method:

$$\begin{aligned}
\hat{\mu} &= \frac{1}{T} \sum_{t=1}^{T} R_t \\
\hat{\Sigma} &= \frac{1}{T} \sum_{t=1}^{T} (R_t - \hat{\mu})(R_t - \hat{\mu})^\top
\end{aligned}$$

We can therefore use the estimates $\hat{\mu}$ and $\hat{\Sigma}$ in place of μ and Σ in mean-variance optimization. This estimation step is very easy. As mentioned by Roncalli (2013), "we could think that the job is complete. However, the story does not end here".

2.2 Evidence of mean-variance instability

Estimating the input parameters of the optimization program necessarily introduces estimation errors and instability in the optimal solution. This stability issue with estimators based on historical figures has been largely studied by academics[3]. Before going into the details of this subject, we propose to illustrate the stability problem of the MVO portfolio with the following example.

Example 1 *We consider a universe of four assets. The expected returns are $\hat{\mu}_1 = 5\%$, $\hat{\mu}_2 = 6\%$, $\hat{\mu}_3 = 7\%$ and $\hat{\mu}_4 = 8\%$ whereas the volatilities are equal to $\hat{\sigma}_1 = 10\%$, $\hat{\sigma}_2 = 12\%$, $\hat{\sigma}_3 = 14\%$ and $\hat{\sigma}_4 = 15\%$. We assume that the correlations are the same and we have $\hat{\rho}_{i,j} = \hat{\rho} = 70\%$.*

We solve the mean-variance problem without constraints using the parameters given in Example 1. The risk tolerance parameter ϕ is calibrated in order to target an ex-ante volatility[4] equal to 10%. In this case the optimal portfolio is $x_1^\star = 23.49\%$, $x_2^\star = 19.57\%$,

[3] See for instance Michaud (1989), Jorion (1992), Broadie (1993) Ledoit and Wolf (2004) or more recently DeMiguel *et al.* (2011).

[4] Let $\sigma^\star$ be the ex-ante volatility. We have:

$$\phi = \frac{\sqrt{\hat{\mu}^\top \hat{\Sigma}^{-1} \hat{\mu}}}{\sigma^\star}$$

$x_3^\star = 16.78\%$ and $x_4^\star = 28.44\%\%$. In Table 1, we indicate how a small perturbation of input parameters changes the optimized solution. For instance, if the volatility of the second asset increases by 3%, the weight on this asset becomes −14.04% instead of 19.57%. If the realized return of the first asset is 6% and not 5%, the optimal weight of the first asset is almost three times larger (63.19% versus 23.49%). As a consequence, the optimized solution is very sensitive to estimation errors.

Table 1: Sensitivity of the MVO portfolio to input parameters

$\hat{\rho}$	70%	80%		80%	
$\hat{\sigma}_2$	12%		15%	15%	
$\hat{\mu}_1$	5%				6%
$x_1^\star$	23.49%	19.43%	36.55%	39.56%	63.19%
$x_2^\star$	19.57%	16.19%	−14.04%	−32.11%	8.14%
$x_3^\star$	16.78%	13.88%	26.11%	28.26%	6.98%
$x_4^\star$	28.44%	32.97%	37.17%	45.87%	18.38%

The stability problem comes from the solution structure. Indeed, the solution involves the inverse of the covariance matrix $\mathcal{I} = \hat{\Sigma}^{-1}$ called the information matrix. The eigenvectors of the two matrices are the same but the eigenvalues of $\mathcal{I}$ are equal to the inverse of the eigenvalues of $\hat{\Sigma}$ (Roncalli, 2013).

Example 2 *We consider our previous example but with another correlation matrix:*

$$\hat{C} = \left(\begin{array}{rrrr} 1.00 & & & \\ 0.50 & 1.00 & & \\ 0.40 & 0.30 & 1.00 & \\ -0.50 & 0.20 & 0.10 & 1.00 \end{array} \right)$$

In Table 2, we consider Example 2 and report the eigenvectors v_j and the eigenvalues λ_j of the covariance and information matrices. Results show that the most important factor[5] of the information matrix is the less important factor of the covariance matrix. However the smallest factors of the covariance matrix are generally considered noise factors because they represent a small part of the total variance. This explains why MVO portfolios are sensitive to input parameters because small changes in the covariance matrix dramatically modify the nature of smallest factors. Despite the simplicity of the mean-variance optimization, the stability of the allocation is then a real problem. In this context, Michaud suggested that mean-variance maximization is in fact 'error maximization':

> "*The unintuitive character of many* optimized *portfolios can be traced to the fact that MV optimizers are, in a fundamental sense,* estimation error maximizers. *Risk and return estimates are inevitably subject to estimation error. MV optimization significantly overweights (underweights) those securities that have large (small) estimated returns, negative (positive) correlations and small (large) variances. These securities are, of course, the ones most likely to have large estimation errors*" (Michaud, 1989, page 33).

In a dynamic framework, estimation errors can then dramatically change the weights leading to high turnover and/or high transaction costs. Moreover, the diversifiable risk is supposed

[5] The j^{th} factor is represented by the eigenvector v_j and the importance of the factor is given by the eigenvalue λ_j.

to be decreased thanks to the optimization that can be underestimated. Aware from these problems, academics and practitioners have developed techniques to reduce the impact of estimation errors.

Table 2: Eigendecomposition of the covariance and information matrices$^{(*)}$ (in %)

v_j	Covariance matrix $\hat{\Sigma}$				Information matrix $\mathcal{I}$			
	v_1	v_2	v_3	v_4	v_1	v_2	v_3	v_4
1	33.68	44.44	−22.21	79.99	79.99	−22.21	44.44	33.68
2	54.04	−0.79	−72.62	−42.48	−42.48	−72.62	−0.79	54.04
3	73.38	8.93	64.94	−17.83	−17.83	64.94	8.93	73.38
4	23.66	−89.13	−3.92	38.47	38.47	−3.92	−89.13	23.66
λ_j	2.66	2.61	1.19	0.20	510.79	83.88	38.37	37.65

(*) The eigenvalues of the information matrix are not expressed in %, but as decimals.

Remark 2 *In Section 3.3.1, we will see how to interpret the eigenvectors and the eigenvalues of the covariance matrix in the MVO framework.*

2.3 Input parameters versus estimation errors

After estimating the input parameters, the optimization is done as if these quantities were perfectly certain, implying that estimation errors are introduced into the allocation process. Various solutions exist to stabilize the optimization from the simplest to the most complicated, but we generally distinguish two ways to regularize the solution.

The first one consists in reducing the estimation errors of the input parameters thanks to econometric methods. For instance, Michaud (1998) uses the resampling approach to reduce the impact of noise estimation. Ledoit and Wolf (2003) propose to replace the covariance estimator by a shrinkage version whereas Laloux *et al.* (1999) clean the covariance matrix thanks to the random matrix theory. Another route is chosen by Black and Litterman (1992), who suggest combining manager views and market equilibrium to modify the expected returns[6].

The second way is to directly shrink the portfolio weights using weight bounds, penalization of the objective function or regularization of input parameters. Jagannathan and Ma (2003) show that imposing constraints on the mean-variance optimization can be interpreted as a modification of the covariance matrix. In particular, lower bounds (resp. upper bounds) decrease (resp. increase) asset return volatilities. Constraints on weights reduce then the degree of freedom of the optimization and the allocation is forced to remain in certain intervals. Instead of using constraints, we can also use other values of input parameters than those estimated with historical figures. For instance, we can consider a diagonal matrix instead of the full covariance matrix or we can use a unique value for the expected returns. This is the case of the equally-weighted (or EW) portfolio, which is the solution for the mean-variance portfolio when $\Sigma = I_n$ and $\mu = \mathbf{1}$. This solution is obtained using

[6] See DeMiguel *et al.* (2011) for a review of shrinkage estimators of the covariance matrix of asset returns and the vector of expected returns.

wrong estimators. However, these estimators have a null variance and minimize the impact of estimation errors on the optimized portfolio.

The correction of estimation errors is such difficult task that several studies tend to show that heuristic allocations perform better than mean-variance allocations in terms of the Sharpe ratio. For example, DeMiguel *et al.* (2009) compare the performances of 14 different portfolio models and the equally-weighted portfolio on different datasets and conclude that sophisticated models are not better than the EW portfolio. More recently, Tu and Zhou (2011) propose to combine the EW portfolio with optimized allocation to outperform naive strategies. In a similar way, Dupleich *et al.* (2012) combine MVO portfolios with different lag windows to remove model uncertainty. By mixing stable noisy portfolios, the authors seek to improve the stability of the allocation. In fact, we will see that most of mixing schemes are equivalent to denoising input parameters.

3 Regularization methods for portfolio optimization

In what follows, we present the most popular techniques used to solve the problem of estimation errors. The first three paragraphs concern weight constraints, resampling methods and shrinkage procedures of the covariance matrix. We then consider the penalization approach of the objective function. Finally, the stability of hedging portfolios based on the information matrix is explained in the last paragraph.

3.1 Using weight constraints

Adding constraints is certainly the first approach that has been used by portfolio managers to regularize optimized portfolios, and it remains today the most frequent method to avoid mean-variance instability.

Let us consider the optimization problem with the normalization constraint:

$$\begin{aligned} x^{\star}(\gamma) &= \arg\max \frac{1}{2} x^{\top} \hat{\Sigma} x - \gamma x^{\top} \hat{\mu} \\ &\text{u.c.} \quad \mathbf{1}^{\top} x = 1 \end{aligned}$$

The constraint $\mathbf{1}^{\top} x = 1$ means that the sum of weights is equal to one. It is easy to show that the optimized portfolio is then:

$$x^{\star}(\gamma; \lambda) = \gamma \hat{\Sigma}^{-1} \tilde{\mu}$$

where $\tilde{\mu} = \hat{\mu} + (\lambda/\gamma) \cdot \mathbf{1}$ and λ is the Lagrange coefficients associated to the constraint. We notice that imposing a portfolio that is fully invested with a leverage equal to exactly one is equivalent to regularize the vector of expected returns. The constraint $\sum_{i=1}^{n} x_i = 1$ is then already a regularization method.

Example 3 *We consider a universe of four assets. The expected returns are $\hat{\mu}_1 = 8\%$, $\hat{\mu}_2 = 9\%$, $\hat{\mu}_3 = 10\%$ and $\hat{\mu}_4 = 8\%$ whereas the volatilities are equal to $\hat{\sigma}_1 = 15\%$, $\hat{\sigma}_2 = 20\%$, $\hat{\sigma}_3 = 25\%$ and $\hat{\sigma}_4 = 30\%$. The correlation matrix is the following:*

$$\hat{C} = \begin{pmatrix} 1.00 & & & \\ 0.10 & 1.00 & & \\ 0.40 & 0.70 & 1.00 & \\ 0.50 & 0.40 & 0.60 & 1.00 \end{pmatrix}$$

If we suppose that $\gamma = 0.5$, we obtain results reported in Table 3. If there is no constraint, the portfolio is highly leveraged. For instance, the weight of the first asset is equal to 207.05%. By adding the simple constraint $\sum_{i=1}^{n} x_i = 1$, the dispersion of optimized weights is smaller[7]. We also notice that the regularized expected returns are lower, because λ is equal to -2.65%.

Table 3: Optimized portfolio with the constraint $\sum_{i=1}^{n} x_i = 1$

	Unconstrained		Constrained	
	μ_i	$x^{\star}(\gamma)$	$\tilde{\mu}_i$	$x^{\star}(\gamma;\lambda)$
1	8.00%	207.05%	2.69%	64.61%
2	9.00%	136.24%	3.69%	42.06%
3	10.00%	−22.75%	4.69%	11.55%
4	8.00%	−32.28%	2.69%	−18.22%

The previous framework may be generalized to other constraints. For instance, Jagannathan and Ma (2003) show that adding a long-only constraint is equivalent to regularizing the covariance matrix. This result also holds for any equality or inequality constraints (Roncalli, 2013). If we consider our previous example and add a long-only constraint, the optimized portfolio is $x_1^{\star} = 52.23\%$, $x_2^{\star} = 42.41\%$, $x_3^{\star} = 1.36\%$ and $x_4^{\star} = 0.00\%$. In this case, the regularized vector of expected returns $\tilde{\mu}$ and the regularized covariance matrix $\tilde{\Sigma}$ are given in Table 4. We notice that the long-only constraint is equivalent to decrease the volatility and the correlation of the fourth asset in order to eliminate its short exposure.

Table 4: Regularized parameters $\tilde{\mu}$ and $\tilde{\Sigma}$

Asset	$\tilde{\mu}_i$	$\tilde{\sigma}_i$	$\tilde{\rho}_{i,j}$			
1	2.83%	15.00%	100.00%			
2	3.83%	20.00%	10.00%	100.00%		
3	4.83%	25.00%	40.00%	70.00%	100.00%	
4	2.83%	26.72%	32.90%	27.49%	53.43%	100.00%

Remark 3 *Portfolio managers generally find the optimal portfolio by sequential steps. They perform the portfolio optimization, analyze the solution to define some regularization constraints, design a new optimization problem by considering these constraints, analyze the new solution and add more satisfying constraints, etc. This step-by-step approach is then very popular, because portfolio managers implicitly regularize the parameters in a coherent way with their expectations for the solution. The drawback may be that the regularized parameters are no longer coherent with the initial parameters. Moreover, the constrained solution is generally overfitted.*

3.2 Resampling methods

Resampling techniques are based on Monte Carlo and bootstrapping methods. Jorion (1992) was the first to apply these techniques to portfolio optimization. The idea is to create more realistic allocation by introducing uncertainty in the decision process of the allocation. For

[7] The weight of the first asset is then equal to 64.61%.

that, we consider a universe of n assets. Let $\hat{\mu}$ and $\hat{\Sigma}$ be the estimates of the expected returns and the covariance matrix of assets returns. The efficient frontier computed with these statistics is an estimation of the true efficient frontier. Michaud (1998) proposed then averaging many realizations of optimized MV solutions to improve out-of-sample performance thanks to the statistical diversification.

The procedure is the following. We generate K samples of asset returns from the original data using Monte Carlo or bootstrap methods:

- Monte Carlo
  The returns are simulated according to a multivariate Gaussian distribution with mean $\hat{\mu}$ and covariance matrix $\hat{\Sigma}$.

- Bootstrap
  The returns are drawn randomly from the original sample with replacement.

We assume that the MV solution is computed for a given value of the risk tolerance. We then calculate the mean $\hat{\mu}_{(k)}$ and the covariance matrix $\hat{\Sigma}_{(k)}$ of the k-th simulated sample. We also calculate the MVO portfolios for a grid of risk tolerance. Finally, we average the weights with respect to the grid and estimate the resampled efficient frontier.

Example 4 *We consider a universe of four assets. The expected returns are* $\hat{\mu}_1 = 5\%$, $\hat{\mu}_2 = 9\%$, $\hat{\mu}_3 = 7\%$ *and* $\hat{\mu}_4 = 6\%$ *whereas the volatilities are equal to* $\hat{\sigma}_1 = 4\%$, $\hat{\sigma}_2 = 15\%$, $\hat{\sigma}_3 = 5\%$ *and* $\hat{\sigma}_4 = 10\%$. *The correlation matrix is the following:*

$$\hat{C} = \begin{pmatrix} 1.00 & & & \\ 0.10 & 1.00 & & \\ 0.40 & 0.20 & 1.00 & \\ -0.10 & -0.10 & -0.20 & 1.00 \end{pmatrix}$$

We illustrate the resampling procedure in Figure 1 by considering Example 4. MVO portfolios are computed under the constraints $\mathbf{1}^\top x = 1$ and $0 \leq x_i \leq 1$. We consider 500 simulated samples and 60 points for the grid. The estimated frontier is calculated with $\hat{\mu}$ and $\hat{\Sigma}$ statistics. The averaged frontier corresponds to the average of the different efficient frontiers obtained for each sample of simulated asset returns. It is different from the resampled frontier, which corresponds to the frontier of resampled portfolios. For instance, we report one optimal resampled portfolio (designed by the red star symbol) which is the average of the 500 resampled portfolios (indicated with the blue cross symbol).

The resampled efficient frontier in Figure 2 is performed with S&P 100 asset returns during the period from January 1, 2011 to December 31, 2011. The resampled frontier is largely below the estimated and averaged efficient frontiers. Moreover, portfolios with high returns are unattainable on the resampled frontier, meaning that these portfolios are extreme points on the estimated efficient frontier and are purely due to estimation noises.

Remark 4 *Resampling techniques have faced some criticisms (Scherer, 2002). The first one concerns the procedure itself, because the resampled portfolio always contains estimation errors since it is computed with the initial parameters* $\hat{\mu}$ *and* $\hat{\Sigma}$. *The second criticism is the lack of theory. Resampling techniques is more an empirical method which seems to correct some biases, because portfolio averaging produces more diversified portfolios. However, they do not solve the robustness question concerning optimized portfolios.*

Figure 1: Simulated resampled efficient frontier (Monte Carlo approach)

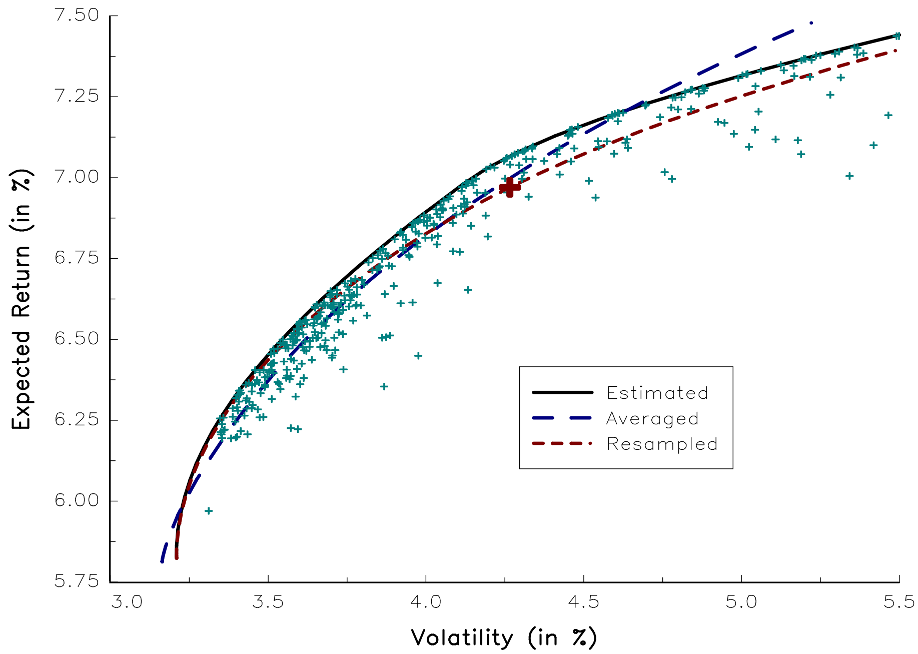
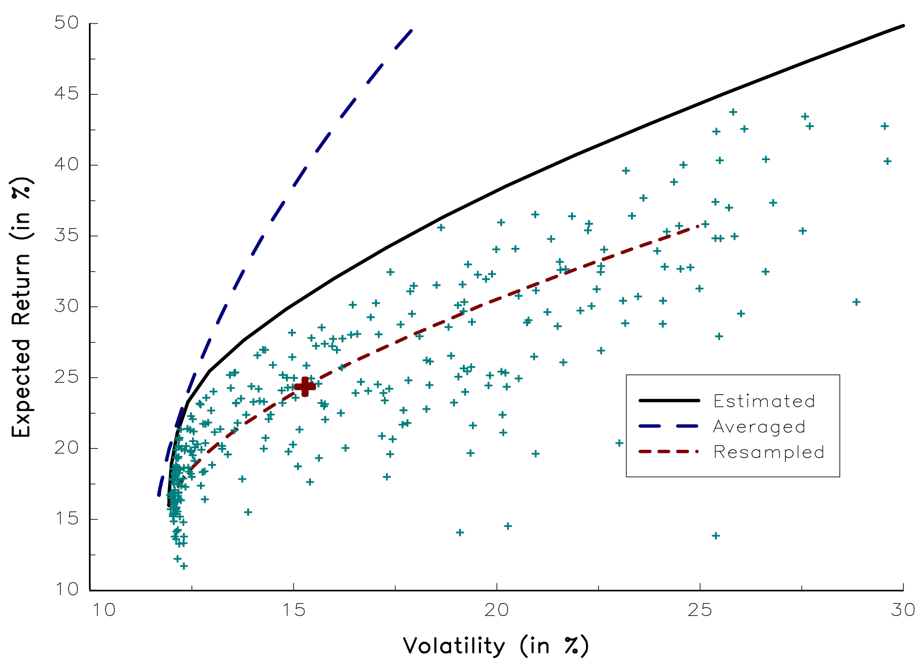


Figure 2: S&P 100 resampled efficient frontier (Bootstrap approach)



3.3 Regularization of the covariance matrix

3.3.1 The eigendecomposition approach

The goal of this method is to reduce the instability of the covariance matrix estimator $\hat{\Sigma}$. For that, we consider the eigendecomposition $\hat{\Sigma} = V \Lambda V^\top$ where $\Lambda = \operatorname{diag}(\lambda_1, \cdots, \lambda_n)$ is the diagonal matrix of the eigenvalues with $\lambda_1 > \lambda_2 > \cdots > \lambda_n$ and V is an orthogonal matrix where each column v_j is an eigenvector. With this decomposition, also known as principal components analysis, we can build endogenous factors $\mathcal{F}_t = \Lambda^{-1/2} V^\top R_t$. In this case, the cleaning process consists in deleting some noise factors $\mathcal{F}_{j,t}$.

Let m be the number of relevant factors. We can then keep the most informative factors, i.e. the factors with the largest eigenvalues. In this case, factors with low eigenvalues are considered as noise factors and we have:

$$m = \max \{ j : \lambda_j \geq (\lambda_1 + \cdots + \lambda_n)/n \}$$

Another solution consists in computing the implicit exposures of the portfolio to these factors. For instance, we can reformulate the MVO portfolio as follows:

$$x^\star = \gamma \hat{\Sigma}^{-1} \hat{\mu} = V \Lambda^{-1/2} \beta$$

where $\beta = \gamma \Lambda^{-1/2} V^\top \hat{\mu}$. If we compute the return of the portfolio's investor, we obtain:

$$R_t(x^\star) = x^{\star\top} R_t = \beta^\top \Lambda^{-1/2} V^\top R_t = \beta^\top \mathcal{F}_t$$

We deduce that β_j is exactly the exposure of the investor to the PCA factor $\mathcal{F}_{j,t}$. We notice that the weights β of the factors in the MVO portfolio are inversely proportional to the square root of the eigenvalues: $\beta_j \propto \sqrt{\lambda_j}$. Thus the optimized portfolio can be strongly exposed to low variance factors, meaning that some noise factors may have a high impact on the MVO solution.

Example 5 *Using the returns of the S&P 100 universe from 2011-2012, we perform the PCA decomposition of the sample correlation matrix. We also compute the implicit exposure β_j to each factor with respect to the MVO portfolio when ϕ is set to 5.*

Table 5: Factor exposures of the MVO portfolio (in %)

Rank	1	7	97	3	73	80
Oil & Gas	11.78	−9.52	−18.89	6.30	12.33	2.98
Basic Materials	4.48	1.05	−11.16	−0.07	−12.61	−5.61
Industrials	16.47	−13.06	2.34	0.68	11.09	19.86
Consumer Goods	11.07	16.41	9.66	6.48	2.63	−6.56
Health Care	9.53	1.98	14.36	8.02	−15.72	4.21
Consumer Services	12.04	−11.21	−16.97	7.92	6.01	8.09
Telecommunications	2.57	−4.65	5.12	−1.54	−15.40	12.56
Utilities	3.62	−11.42	14.88	−4.76	−10.13	−7.28
Financials	16.98	25.37	2.29	−40.87	−4.18	−30.09
Technology	11.47	5.32	4.33	23.36	9.90	−2.76
β_j	−0.16	−2.90	2.87	2.49	2.45	−2.41

Results are reported in Table 5. For each factor, we give the loading coefficients with respects to ICB classification[8]. The second column is the eigenvector with the largest eigenvalue. You can see that it is a proxy of a sector-weighted portfolio[9]. It may be viewed as a market factor. The other reported factors are the top five most important factors of the MVO portfolio in terms of beta exposures. These factors correspond to long-short portfolios of industry sectors. We notice that the factor with the highest beta is ranked 7, whereas the second most important factor corresponds to the factor ranked 97, which is certainly a noise factor. We verify that the beta exposure of the market factor is very small[10]. This example illustrates how some factors can introduce noise in the MVO solution and how an investor can be exposed to non significant factors. A way to reduce this noise is then to set to 0 the weight of these noisy factors.

A last solution consists in using random matrix theory to regularize eigenvalues of the correlation matrix. Thanks to the random matrix theory, Laloux *et al.* (1999) showed that the eigenvalues of the estimated correlation matrix are generally more dispersed than the true ones. A first consequence for the MVO allocation is the overweighing of some assets. Indeed the optimization focuses on some low eigenvalues whereas these eigenvalues were equal to the others in the true correlation matrix. Random matrix theory allows to test if the dispersion of the eigenvalues is significant or just due to noise. As a consequence, regularizing the estimated correlation matrix would be either to delete or equalize the eigenvalues, which are not significant. Laloux *et al.* (1999) studied the estimated correlation matrix of n identical independent asset returns based on T observations and showed that the eigenvalues follow a Marcenko-Pastur (MP) distribution[11]:

$$\rho(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{\max} - \lambda)(\lambda - \lambda_{\min})}}{\lambda}$$

where $Q = T/n$. The maximum and minimum eigenvalues are then given by:

$$\lambda_{\min}^{\max} = \sigma^2 \left(1 \pm \sqrt{1/Q}\right)^2$$

It is therefore difficult to distinguish the true eigenvalues from noisy eigenvalues for a matrix whose eigenvalue distribution looks like the MP distribution. On the other hand, the eigenvalue spectrum outside this distribution could represent real information.

Example 6 *We compute the theoretical distribution of λ for different value of T when $n = 100$. We also simulate the eigenvalue distribution of independent asset returns with $n = 100$ and $T = 260$. We finally consider the eigenvalue distribution of S&P 100 asset returns for the year 2011.*

In the first panel in Figure 3, we report on the Marcenko-Pastur distribution of the eigenvalues. In the second panel, we compare the histogram of simulated independent asset returns (red bars) and the theoretical MP distribution (blue line). The last panel corresponds to the eigenvalues of the correlation matrix in the case of the S&P 100 universe. For that, we remove the first eigenvalue which represents 60% of the total variance. If we consider the 99 remaining eigenvalues, we observe that their histogram is close to the MP distribution,

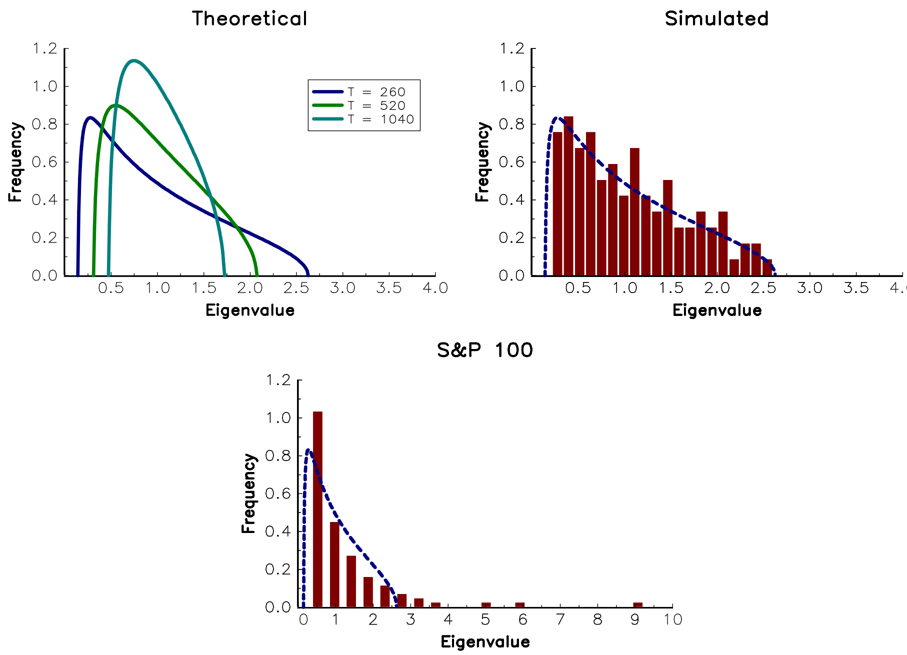
[8]The ICB repartition of the 100 stocks is the following: Energy (11), Basic Materials (4), Industrials (15), Consumer Goods (12), Health Care (10), Consumer Services (13), Telecommunications (3), Utilities (4), Financials (16) and Technology (12).

[9]The weight of the sector is closed to the frequency of stocks belonging to it.

[10]It is equal to -0.16%.

[11]See Marcenko and Pastur (1967).

Figure 3: Eigenvalue distribution



except for seven eigenvalues which are outside the dashed blue line. Denoising the correlation matrix can then be performed by replacing all the eigenvalues under the dashed blue line by their mean.

3.3.2 The shrinkage approach

This method was popularized by Ledoit and Wolf (2003). They propose do define the shrinkage estimator of the covariance matrix as a combination of the sample estimator of the covariance matrix $\hat{\Sigma}$ and a target covariance matrix $\hat{\Phi}$:

$$\tilde{\Sigma}_\alpha = \alpha\hat{\Phi} + (1-\alpha)\,\hat{\Sigma}$$

where α is a constant between 0 and 1. We know that $\hat{\Sigma}$ is a non-biased estimator, but its convergence is slow. The underlying idea is then to combine it with a biased estimator $\hat{\Phi}$, but which converges faster. As a result, the mean squared error of the estimator is reduced. This approach is very close to the principle of bias and variance trade-off well known in regression analysis (James and Hastie, 1997). Ledoit and Wolf (2003) use the bias-variance decomposition with respect to the Frobenius norm to propose an optimal shrinkage parameter $\alpha^\star$. The loss function considered by Ledoit and Wolf is the following:

$$L(\alpha) = \left\|\alpha\hat{\Phi} + (1-\alpha)\,\hat{\Sigma} - \Sigma\right\|^2$$

By solving the minimization problem $\alpha^\star = \arg\min \mathbb{E}\left[L(\alpha)\right]$, they give an analytical expression of $\alpha^\star$.

Ledoit and Wolf (2003) consider the single-factor model of Sharpe (1963). In this case, the vector of asset returns R_t can be written as a function of the market return $R_{m,t}$ and uncorrelated Gaussian residuals $\varepsilon_t \sim \mathcal{N}(\mathbf{0}, D)$:

$$R_t = \beta R_{m,t} + \varepsilon_t$$

where β is the vector of market betas, σ_m is the volatility of the market portfolio and $D = \operatorname{diag}\left(\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_n^2\right)$ is the covariance of specific risks. The covariance matrix $\hat{\Phi}$ of the single-factor model is then:

$$\hat{\Phi} = \sigma_m^2 \beta \beta^\top + D$$

Assuming that the first eigenvector of $\hat{\Sigma}$ is the market factor, we obtain[12]:

$$\tilde{\Sigma}_\alpha \simeq \lambda_1 v_1 v_1^\top + \sum_{i=2}^{n} \left((1-\alpha)\lambda_i + \alpha \tilde{\sigma}^2\right) v_i v_i^\top$$

The expression $(1-\alpha)\lambda_i + \alpha\tilde{\sigma}^2$ shows that shrinking toward the single-factor matrix is equivalent to modifying the distribution of eigenvalues. The highest eigenvalue is unchanged whereas the other eigenvalues are forced to be closer to specific risks. Others models can be considered like the constant correlation matrix (Ledoit and Wolf, 2004), but the result of the shrinkage approach is always to reduce the dispersion of eigenvalues.

3.4 Penalization methods

The idea of using penalizations comes from the regularization problem of linear regressions. These techniques have been largely used in machine learning in order to improve out-of-sample forecasting (Tibshirani, 1996; Zou and Hastie, 2005). Since mean-variance optimization is related to linear regression (Scherer, 2007), regularizations may improve the performance of MVO portfolios. For instance, DeMiguel *et al.* (2010) consider the following norm-constrained problem:

$$\begin{aligned} x^\star(\lambda) &= \arg\min \frac{1}{2} x^\top \hat{\Sigma} x + \lambda \|x\| \\ &\text{u.c.} \quad \mathbf{1}^\top x = 1 \end{aligned}$$

where $\|x\|$ is the norm of the portfolio x. In particular they proved that the solution of the L_1 norm-constrained MV problem is the same as the short-sale constrained minimum-variance portfolio analyzed in Jagannathan and Ma (2003). They also demonstrate that using the L_2 norm is equivalent to combine MV and EW portfolios.

3.4.1 The L_1 constrained portfolio

The L_1 norm or the Lasso approach is one of the most famous regularization procedures. The penalty consists to constrain the sum of the absolute values of the weights. We have[13]:

$$x^\star(\gamma, \lambda) = \arg\min \frac{1}{2} x^\top \hat{\Sigma} x - \gamma x^\top \hat{\mu} + \lambda \|x\| \tag{1}$$

The L_1 penalty has useful properties. It improves the sparsity and thus the selection of assets in large portfolio. Moreover, it stabilizes the problem by imposing size restriction on the weights. Even there is no closed solution of Equation (1), it can be easily solved with QP

[12] See Appendix A.1 for computational details. We also assume that the idiosyncratic volatilities are equal ($\tilde{\sigma}_1 = \ldots = \tilde{\sigma}_n = \tilde{\sigma}$).

[13] The L_1 norm is defined as follows: $\|x\| = \sum_{i=1}^{n} |x_i|$. It may be interpreted as the portfolio leverage.

algorithm. If the covariance matrix is the identity, we obtain an analytical formula which gives insight on the effect on the L_1 norm. The solution is[14]:

$$x^\star(\gamma,\lambda) = \operatorname{sgn}(\hat{\mu}) \cdot (\gamma |\hat{\mu}| - \lambda)^+$$

The L_1 norm corresponds then to a soft-thresholding operator of the expected return.

The L_1 penalty is also well adapted to portfolio optimization under transaction or liquidity costs (Scherer, 2007). Let c be the vector of transaction costs and x_0 the initial portfolio. The transaction cost paid by the investor is $c^\top |x^\star - x_0|$ and may be easily introduced into the mean-variance optimization. Another way to use the L_1 norm is to perform asset selection. The investor may then choose the parameter λ which corresponds to the given number m of selected assets.

Example 7 *We consider the asset returns of the S&P 100 universe for the period January 2011 – December 2011. We compute the regularized L_1 MVO portfolio for different values of λ.*

Results are reported in Figure 4. In the first panel, we indicate the number of selected stocks. The optimized value of the utility function (or the ex-ante Sharpe ratio) is given in the second panel. We also report the weight evolution of the consumer services stocks. Finally, we indicate the leverage $\sum_{i=1}^n |x_i|$ of the portfolio in the last panel. This example illustrates the sparsity property when λ increases. We also notice the impact of λ in the leverage of the portfolio. For instance, if $\lambda = 0.2\%$, the leverage is divided by a factor larger than six whereas the decrease of the utility function is equal to 28%. As a result, we may obtain more sparse portfolios with limited impacts on the ex-ante Sharpe ratio.

3.4.2 The L_2 constrained portfolio

The L_2 constrained MVO problem is defined as follows:

$$\begin{aligned} x^\star(\gamma,\lambda) &= \arg\min \frac{1}{2} x^\top \hat{\Sigma} x - \gamma x^\top \hat{\mu} + \frac{1}{2}\lambda x^\top x \\ &= \arg\min \frac{1}{2} x^\top \left(\hat{\Sigma} + \lambda I_n\right) x - \gamma x^\top \hat{\mu} \end{aligned}$$

$x^\star(\gamma,\lambda)$ is then a MVO portfolio with a modified covariance matrix $\tilde{\Sigma} = \hat{\Sigma} + \lambda I_n$. Imposing the L_2 constraint is equivalent to adding the same amount λ to the diagonal elements of the covariance matrix. This approach is therefore very close to the shrinkage method of Ledoit and Wolf (2003).

Remark 5 *The L_2 constraint may be viewed as an eigenvalue shrinkage method. Indeed, we have $\hat{\Sigma} = V \Lambda V^\top$ and $\tilde{\Sigma} = V(\Lambda + \lambda I_n) V^\top$ because $VV^\top = I_n$. The parameter λ is thus useful to stabilize the small eigenvalues of the covariance matrix.*

[14] We notice that $\operatorname{sgn}(x^\star) = \operatorname{sgn}(\hat{\mu})$. The first order condition is $x - \gamma\hat{\mu} + \lambda \operatorname{sgn}(\hat{\mu}) = \mathbf{0}$. We deduce that:

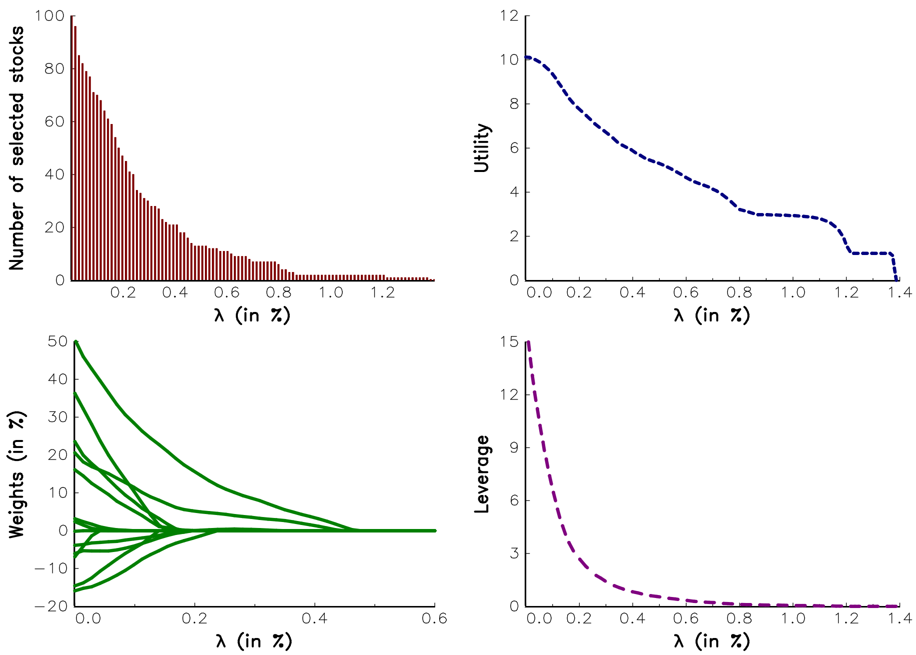
- If $\hat{\mu}_i \geq 0$, the first order condition becomes $x_i - \gamma\hat{\mu}_i + \lambda = 0$ and we have:

$$\begin{aligned} x_i^\star &= \gamma\hat{\mu}_i - \lambda \\ &= \operatorname{sgn}(\hat{\mu}_i) \cdot (\gamma |\hat{\mu}_i| - \lambda)^+ \end{aligned}$$

- If $\hat{\mu}_i < 0$, the first order condition becomes $x_i - \gamma\hat{\mu}_i - \lambda = 0$ and we have:

$$\begin{aligned} x_i^\star &= \gamma\hat{\mu}_i + \lambda \\ &= \operatorname{sgn}(\hat{\mu}_i) \cdot (\gamma |\hat{\mu}_i| - \lambda)^+ \end{aligned}$$

Figure 4: Illustration of the L_1 norm-constrained portfolio optimization



The solution can be written as a linear combination of the MVO solution $x^{\star}(\gamma)$:

$$\begin{aligned} x^{\star}(\gamma,\lambda) &= \gamma\left(\hat{\Sigma}+\lambda I_n\right)^{-1}\hat{\mu} \\ &= \left(I_n+\lambda\hat{\Sigma}^{-1}\right)^{-1}x^{\star}(\gamma) \end{aligned}$$

Using the eigendecomposition $\hat{\Sigma}=V\Lambda V^{\top}$, the solution can be expressed in a simple form:

$$\begin{aligned} x^{\star}(\gamma,\lambda) &= \left(VV^{\top}+\lambda V\Lambda^{-1}V^{\top}\right)^{-1}x^{\star}(\gamma) \\ &= V\tilde{\Lambda}V^{\top}x^{\star}(\gamma) \end{aligned}$$

where $\tilde{\Lambda}$ is a diagonal matrix with elements $\tilde{\Lambda}_j=\Lambda_j/(\Lambda_j+\lambda)$. We notice that the weights are equal to 0 when $\lambda=+\infty$.

Instead of using the identity matrix, we can consider a general matrix A to define the L_2 norm:

$$\begin{aligned} x^{\star}(\gamma,\lambda) &= \arg\min\frac{1}{2}x^{\top}\hat{\Sigma}x-\gamma x^{\top}\hat{\mu}+\frac{1}{2}\lambda x^{\top}Ax \\ &= \arg\min\frac{1}{2}x^{\top}\left(\hat{\Sigma}+\lambda A\right)x-\gamma x^{\top}\hat{\mu} \end{aligned}$$

The solution is then:

$$\begin{aligned} x^{\star}(\gamma,\lambda) &= \gamma\left(\hat{\Sigma}+\lambda A\right)^{-1}\hat{\mu} \\ &= \left(I_n+\lambda\hat{\Sigma}^{-1}A\right)^{-1}x^{\star}(\gamma) \end{aligned}$$

In the case of L_2 identity constraint, the covariance matrix is shrunken toward the identity matrix. If A is the diagonal matrix of asset variances ($A = \upsilon$), the shrinkage is based on the correlation matrix[15]. This approach is sometimes used by portfolio managers when they reduce the correlations even if they don't realize it. Indeed, we have:

$$\begin{aligned} x^\star(\gamma, \lambda) &= \gamma \left(\hat{\Sigma} + \lambda \upsilon\right)^{-1} \hat{\mu} \\ &= \frac{\gamma}{1+\lambda} \left(\eta \hat{\Sigma} + (1-\eta)\,\upsilon\right)^{-1} \hat{\mu} \end{aligned}$$

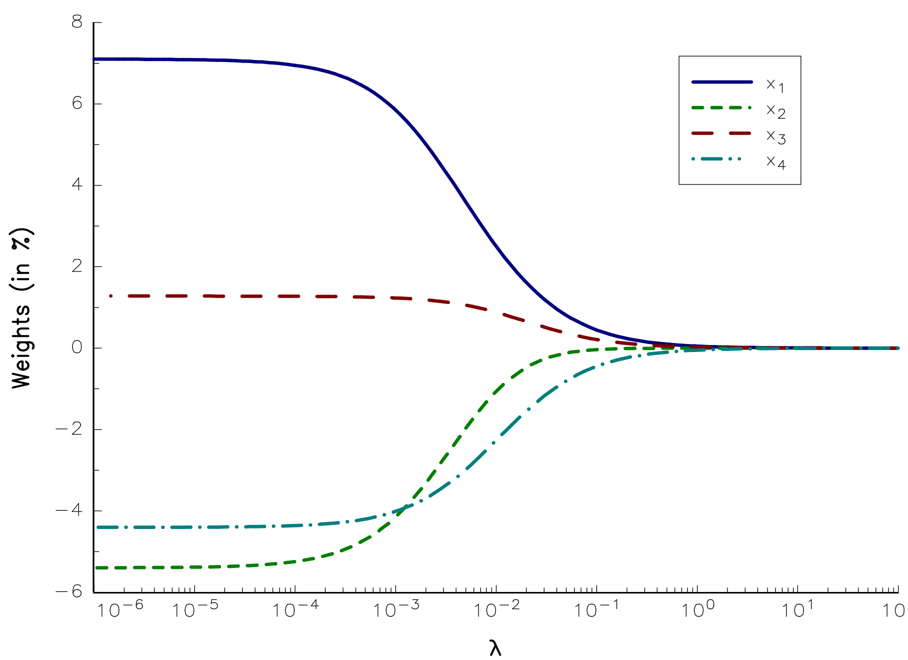
with $\eta = (1+\lambda)^{-1}$. In this case, the solution $x^\star(\gamma, \lambda)$ is an optimized portfolio where we keep a percentage η of the correlations.

Example 8 *To illustrate the L_2 approach, we consider a universe of four assets. The correlation matrix is:*

$$\hat{C} = \begin{pmatrix} 1.00 & & & \\ 0.60 & 1.00 & & \\ 0.20 & 0.20 & 1.00 & \\ -0.20 & -0.20 & -0.20 & 1.00 \end{pmatrix}$$

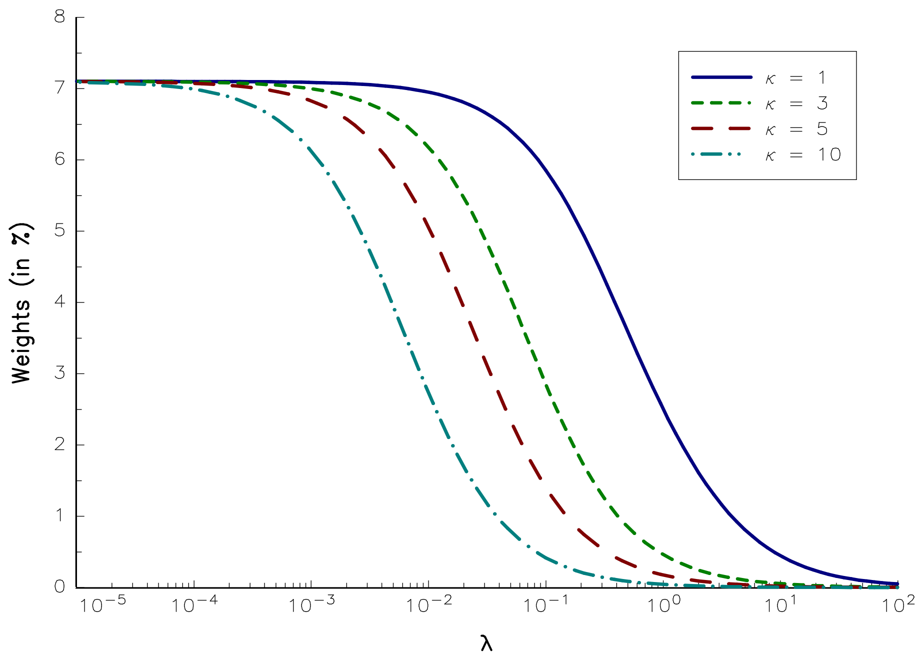
The expected returns are 10%, 0%, 5% and 10% whereas the volatilities are the same and are equal to 10%.

Figure 5: Weights with respect to λ



[15] See Appendix A.2.1.

Figure 6: Effect of the penalty matrix



We assume that $\gamma = 0.5\%$. In Figure 5, we report the evolution of the weights when $A = I_n$. We verify that the solution is the MVO portfolio if $\lambda = 0$ and tends to 0 if λ increases. In Figure 6, we consider that $A = \operatorname{diag}\left(\kappa\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_3^2, \hat{\sigma}_4^2\right)$. We observe the impact on the weight x_1 of the first asset when the uncertainty κ on this asset increases.

The L_2 portfolio optimization can also be used when investors target a portfolio x_0:

$$x^{\star}(\gamma, \lambda) = \arg\min \frac{1}{2} x^{\top} \hat{\Sigma} x - \gamma x^{\top} \hat{\mu} + \frac{1}{2} \lambda (x - x_0)^{\top} A (x - x_0) \tag{2}$$

The parameter λ controls the distance between the MVO portfolio and the target portfolio. For instance, the target portfolio could be an heuristic allocation like the EW, MV or ERC portfolio (Roncalli, 2013) or it could be the actual portfolio in order to limit the turnover. In this case, we interpret λ as risk aversion with respect to the MVO portfolio. We notice that the analytical solution is:

$$x^{\star}(\gamma, \lambda) = \left(\hat{\Sigma} + \lambda A\right)^{-1} (\gamma \hat{\mu} + \lambda A x_0)$$

If $A = I_n$, the optimal portfolio becomes:

$$\begin{aligned} x^{\star}(\gamma, \lambda) &= \left(\hat{\Sigma} + \lambda I_n\right)^{-1} (\gamma \hat{\mu} + \lambda x_0) \\ &= \gamma \tilde{\Sigma}^{-1} \tilde{\mu} \end{aligned}$$

where $\tilde{\Sigma} = \hat{\Sigma} + \lambda I_n$ and $\tilde{\mu} = \hat{\mu} + (\lambda / \gamma) x_0$. This approach corresponds to a double shrinkage of the covariance matrix $\hat{\Sigma}$ and the vector of expected returns $\hat{\mu}$ (Candelon *et al.*, 2012). We can also reformulate the solution as follows[16]:

$$x^{\star}(\gamma, \lambda) = B x^{\star}(\gamma) + (I_n - B) x_0$$

[16] See Appendix A.2.2.

where $B = \left(I_n + \lambda \hat{\Sigma}^{-1}\right)^{-1}$. The optimal portfolio is then a linear combination between the MVO portfolio and the target portfolio, and coincides with classical allocation policy. For example, when an investor considers a 50/50 allocation policy, B is equal to $I_n/2$ and we obtain:

$$x^{\star}(\gamma, \lambda) = \frac{1}{2} x^{\star}(\gamma) + \frac{1}{2} x_0$$

Example 9 *We consider a universe of three assets. The expected returns are* 5%, 6% *and* 7% *whereas the volatilities are* 10%, 15% *and* 20%. *The correlation matrix is:*

$$\hat{C} = \left(\begin{array}{ccc} 1.00 & & \\ 0.50 & 1.00 & \\ 0.20 & -0.30 & 1.00 \end{array}\right)$$

The risk aversion parameter γ *is set to* 30%. *We assume that the target portfolio is the EW portfolio.*

Table 6: L_2 portfolio with a target allocation

asset	x_0	$x^{\star}(\gamma)$	$x^{\star}(\gamma, \lambda)$ $\lambda = 0.01$	$\lambda = 0.10$	$\lambda = 1.00$
1	33.33	58.13	54.55	39.58	34.10
2	33.33	87.14	68.75	42.45	34.41
3	33.33	66.29	56.68	40.41	34.24

In Table 6, we report the solution for different values of λ. When λ is small, the diagonal elements of B are high and the MVO portfolio $x^{\star}(\gamma)$ dominates the target portfolio x_0. For instance, if $\lambda = 1\%$, we obtain:

$$B = \left(\begin{array}{ccc} 0.43 & 0.15 & 0.07 \\ 0.15 & 0.64 & -0.08 \\ 0.07 & -0.08 & 0.78 \end{array}\right)$$

If we are interested to reduce the turnover, we can use a time-varying regularization:

$$x_t^{\star}(\gamma, \lambda) = \arg\min \frac{1}{2} x_t^{\top} \hat{\Sigma}_t x_t - \gamma x_t^{\top} \hat{\mu}_t + \frac{1}{2} \lambda \left(x_t - x_{t-1}\right)^{\top} A \left(x_t - x_{t-1}\right) \qquad (3)$$

where $t-1$ and t are two successive rebalancing dates and x_{t-1} is the previous allocation. The analytical solution is:

$$\begin{array}{rcl} x_t^{\star}(\gamma, \lambda) & = & \gamma \left(\hat{\Sigma}_t + \lambda A\right)^{-1} \hat{\mu}_t + \lambda \left(\hat{\Sigma}_t + \lambda A\right)^{-1} A x_{t-1} \\ & = & B_t x_t^{\star}(\gamma) + \left(I_n - B_t\right) x_{t-1} \end{array}$$

with $B_t = \left(I_n + \lambda \hat{\Sigma}_t^{-1} A\right)^{-1}$. If we assume that $x_{t-1} = x_{t-1}^{\star}(\gamma, \lambda)$, it follows that the current allocation is a moving average of past unconstrained MVO portfolios:

$$x_t^{\star}(\gamma, \lambda) = B_t x_t^{\star}(\gamma) + \sum_{i=1}^{t} \prod_{j=0}^{i-1} \left(I_n - B_{t-j}\right) B_{t-i} x_{t-i}^{\star}(\gamma)$$

Remark 6 *Suppose that* $\hat{\Sigma}_t = \hat{\Sigma}_{t-1}$ *and* $A = \operatorname{diag}\left(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2\right)$. *If asset returns are not correlated, we obtain:*

$$\begin{aligned} x_{i,t}^{\star}(\gamma, \lambda) &= \frac{\gamma}{\hat{\sigma}_i^2 + \lambda \hat{\sigma}_i^2} \hat{\mu}_{i,t} + \frac{\lambda \hat{\sigma}_i^2}{\hat{\sigma}_i^2 + \lambda \hat{\sigma}_i^2} x_{i,t-1}^{\star}(\gamma, \lambda) \\ &= \alpha x_{i,t}^{\star}(\gamma) + (1-\alpha) x_{i,t-1}^{\star}(\gamma, \lambda) \end{aligned}$$

where $\alpha = 1/(1+\lambda)$. *The solution is an exponentially weighted moving average filter. Calibrating* λ *is then equivalent to choosing the holding period to turn the portfolio.*

3.5 Information matrix and hedging portfolios

The previous methods are focused on the covariance matrix. However, the important quantity in mean-variance optimization is the information matrix $\mathcal{I} = \hat{\Sigma}^{-1}$, i.e. the inverse of the covariance matrix (Scherer, 2007; Roncalli, 2013). Stevens (1998) gives a new interpretation of the information matrix using the following regression framework:

$$R_{i,t} = \beta_0 + \beta_i^{\top} R_t^{(-i)} + \varepsilon_{i,t} \tag{4}$$

where $R_t^{(-i)}$ denotes the vector of asset returns R_t excluding the i^{th} asset and $\varepsilon_{i,t} \sim \mathcal{N}(0, s_i^2)$. Let R_i^2 be the R-squared of the linear regression (4) and $\hat{\beta}$ be the matrix of OLS coefficients with rows $\hat{\beta}_i^{\top}$. Stevens (1998) shows that the diagonal elements of the information matrix are given by:

$$\mathcal{I}_{i,i} = \frac{1}{\hat{\sigma}_i^2 \left(1 - R_i^2\right)}$$

whereas the off-diagonal elements are:

$$\mathcal{I}_{i,j} = -\frac{\hat{\beta}_{i,j}}{\hat{\sigma}_i^2 \left(1 - R_i^2\right)} = -\frac{\hat{\beta}_{j,i}}{\hat{\sigma}_j^2 \left(1 - R_j^2\right)}$$

Using this expression of $\mathcal{I}$, we obtain a new formula of the MVO portfolio:

$$x_i^{\star}(\gamma) = \gamma \frac{\hat{\mu}_i - \hat{\beta}_i^{\top} \hat{\mu}^{(-i)}}{\hat{\sigma}_i^2 \left(1 - R_i^2\right)}$$

Scherer (2007) interprets $\hat{\mu}_i - \hat{\beta}_i^{\top} \hat{\mu}^{(-i)}$ as the excess return after regression hedging and $\hat{\sigma}_i^2 \left(1 - R_i^2\right)$ as the non-hedging risk. We remind that $R_i^2 = 1 - \hat{s}_i^2 / \hat{\sigma}_i^2$. We finally obtain:

$$x_i^{\star}(\gamma) = \gamma \frac{\hat{\mu}_i - \hat{\beta}_i^{\top} \hat{\mu}^{(-i)}}{\hat{s}_i^2}$$

From this equation, we deduce the following conclusions:

1. The better the hedge, the higher the exposure. This is why highly correlated assets produces unstable MVO portfolios.
2. The long-short position is defined by the sign of $\hat{\mu}_i - \hat{\beta}_i^{\top} \hat{\mu}^{(-i)}$. If the expected return of the asset is lower than the conditional expected return of the hedging portfolio, the weight is negative.

It has been shown that the linear regression can be improved using norm constraints (Hastie *et al.*, 2009). For example we can use the L_2 regression to improve the predictive power of the hedging relationships. We can also estimate the hedging portfolios with the L_1 penalty.

Example 10 *We consider a universe of four assets. The expected returns are $\hat{\mu}_1 = 7\%$, $\hat{\mu}_2 = 8\%$, $\hat{\mu}_3 = 9\%$ and $\hat{\mu}_4 = 10\%$ whereas the volatilities are equal to $\hat{\sigma}_1 = 15\%$, $\hat{\sigma}_2 = 18\%$, $\hat{\sigma}_3 = 20\%$ and $\hat{\sigma}_4 = 25\%$. The correlation matrix is the following:*

$$\hat{C} = \begin{pmatrix} 1.00 & & & \\ 0.50 & 1.00 & & \\ 0.50 & 0.50 & 1.00 & \\ 0.60 & 0.50 & 0.40 & 1.00 \end{pmatrix}$$

In Table 7, we have reported the results of the hedging portfolios. The OLS coefficients $\hat{\beta}_i$, the coefficient of determination R_i^2 and the standard error $\hat{s}_i$ of residuals are computed thanks to the formulas given in Appendix A.4. We also have computed the conditional expected return[17] $\bar{\mu}_i = \hat{\mu}_i - \hat{\beta}_i^\top \hat{\mu}^{(-i)}$. We can then deduce the corresponding information matrix $\mathcal{I}$ and the MVO portfolio $x^\star$ for $\gamma = 0.5$. We finally obtain a very well balanced allocation, because the weights range between 19.28% and 69.80%. Let us now change the value of the correlation between the third and fourth assets. If $\rho_{3,4} = 95\%$, we obtain results given in Table 8. In this case, the story is different, because the optimized portfolio is not well balanced. Indeed, because two assets are strongly correlated, some hedging relationships present high value of R^2. The information matrix is then very sensitive to these hedging portfolios. This explains that the weights are now in the range between -168.70% and 239.34%!

Table 7: Hedging portfolios when $\rho_{3,4} = 40\%$

Asset	$\hat{\beta}_i$				R_i^2	$\hat{s}_i$	$\bar{\mu}_i$	$x^\star$
1		0.139	0.187	0.250	45.83%	11.04%	1.70%	69.80%
2	0.230		0.268	0.191	37.77%	14.20%	2.06%	51.18%
3	0.409	0.354		0.045	33.52%	16.31%	2.85%	53.66%
4	0.750	0.347	0.063		41.50%	19.12%	1.41%	19.28%

Table 8: Hedging portfolios when $\rho_{3,4} = 95\%$

Asset	$\hat{\beta}_i$				R_i^2	$\hat{s}_i$	$\bar{\mu}_i$	$x^\star$
1		0.244	−0.595	0.724	47.41%	10.88%	3.16%	133.45%
2	0.443		0.470	−0.157	33.70%	14.66%	2.23%	52.01%
3	−0.174	0.076		0.795	91.34%	5.89%	1.66%	239.34%
4	0.292	−0.035	1.094		92.38%	6.90%	−1.61%	−168.67%

4 Some applications

In this section, we look at three applications which are directly linked to the previous framework. The first application concerns the relationship between the MVO portfolio and the principal portfolios derived from PCA analysis. The second application shows the usefulness of the L_2 covriance matrix regularization. We finally illustrate how the Lasso approach may improve the robustness of the information matrix and hedging portfolios in the third application.

[17] We note that $\bar{\mu}_i$ is also equal to the intercept $\hat{\beta}_0$ of the linear regression.

4.1 Principal portfolios

We first consider the problem of mean-variance optimization in a multi-assets universe. We have shown that this portfolio has implicit exposition to arbitrage and risk factors that we call *principal* portfolios (Meucci, 2009). The universe is composed of ten indices, four developed market equity indexes, one emerging market equity index, two bond indexes, two currency indexes and one commodity index: S&P 500 index, Eurostoxx 50 index, Topix index, Russell 2000 index, MSCI EM index, Merrill Lynch US High Yield index, JP Morgan Emerging Bonds index, EUR/USD and JPY/USD exchange rates and S&P GSCI Commodity index.

To better understand the importance of principal portfolios, we look at the mean-variance optimization at the end of 2006. The covariance matrix is estimated using historical daily returns from January 2006 to December 2006. We then assume that this portfolio is held until the end of 2007. For the purpose of the study, we will consider perfect views of the market. The expected returns are therefore the realized returns of each asset from January 2007 to December 2007. We report in Table 9 the expected return, the volatility and the Sharpe ratio of each asset. The mean-variance optimized portfolio is computed under a 10% volatility constraint[18]. The optimized weights are given in the last column in Table 9.

Table 9: Statistics and MVO portfolio at the end of 2006

Asset	$\hat{\mu}_i$	$\hat{\sigma}_i$	SR_i	$x_i^\star$
SPX	1.27%	10.04%	0.13	65.57%
SX5E	6.24%	14.65%	0.43	−1.97%
TPX	−10.60%	18.85%	−0.56	−24.39%
RTY	−4.84%	17.37%	−0.28	−49.54%
MSCI EM	30.15%	18.11%	1.66	55.60%
US HY	−3.07%	1.69%	−1.82	−489.99%
EMBI	0.50%	4.08%	0.12	29.26%
EUR/USD	9.27%	7.72%	1.20	41.91%
JPY/USD	2.04%	8.28%	0.25	−20.96%
GSCI	24.76%	21.17%	1.17	2.62%

Let us now study the decomposition of the MVO portfolio by its principal portfolios. Their weights are given in Table 10. By construction, each principal portfolio is independent from the others and is composed by all the assets (Meucci, 2009). The MVO portfolio is then a combination of these principal portfolios. It follows that the optimal weight x_i^* of an asset i is the sum of the exposure β_j of the principal portfolio $\mathcal{F}_j$ in the MVO portfolio multiplied by the weight $w_{i,j}$ of the asset i in the portfolio $\mathcal{F}_j$:

$$x_i^* = \sum_{j=1}^{10} \beta_j \cdot w_{i,j}$$

We have also reported the expected return μ_j and the ex-ante volatility σ_j associated to each principal portfolio $\mathcal{F}_j$. The first portfolio $\mathcal{F}_1$ is the riskier portfolio. We notice that it is a long-only portfolio and it has the profile of a risk weighted portfolio. Assets with a lower volatility have then a lower weight. As a result, the other principal portfolios can

[18] In this case, γ is equal to 2.39%.

be considered as neutral risk weighted portfolios, because they are uncorrelated to the first principal portfolio. Nevertheless, if we consider the exposures β_j, we observe that the first principal portfolio is underweighted compared to the other principal portfolios.

Table 10: Decomposition of the MVO portfolio at the end of 2006

Asset	$\mathcal{F}_1$	$\mathcal{F}_2$	$\mathcal{F}_3$	$\mathcal{F}_4$	$\mathcal{F}_5$	$\mathcal{F}_6$	$\mathcal{F}_7$	$\mathcal{F}_8$	$\mathcal{F}_9$	$\mathcal{F}_{10}$
SPX	22.52	-22.39	26.56	-18.43	-12.03	-6.50	-8.73	86.49	14.24	-0.37
SX5E	37.59	-20.94	14.89	17.05	78.50	-36.98	4.23	-8.30	4.37	0.56
TPX	42.23	-0.13	-74.05	-46.21	-7.01	-23.06	2.26	-0.11	-3.09	1.17
RTY	41.56	-37.56	47.66	-36.01	-32.97	6.93	8.81	-45.49	-3.28	-0.78
MSCI EM	56.86	-2.69	-19.41	57.13	-11.10	53.91	-4.38	3.44	7.63	0.76
US HY	2.41	-1.06	-0.65	3.81	-1.15	-2.77	-0.51	2.98	-22.49	-97.23
EMBI	7.72	-2.22	4.57	9.42	-2.96	-5.76	-11.15	11.75	-94.80	23.10
EUR/USD	5.92	5.36	0.88	33.88	-32.34	-42.44	76.71	6.88	-0.12	2.84
JPY/USD	7.05	-0.80	1.48	34.19	-37.01	-56.97	-61.52	-13.21	14.31	0.18
GSCI	34.45	87.22	30.28	-15.93	2.77	-4.03	-3.14	0.51	0.97	-1.03
β_j	7.11	12.11	6.14	39.16	1.50	34.59	31.02	75.86	95.32	484.58
μ_j	22.47	21.53	8.61	24.52	0.69	9.94	3.21	4.29	3.98	3.29
σ_j	27.51	20.63	18.32	12.25	10.48	8.30	4.98	3.68	3.16	1.27
SR_j	0.82	1.04	0.47	2.00	0.66	1.20	0.64	1.17	1.26	2.58

Let us consider the other principal portfolios. For instance, the principal portfolio $\mathcal{F}_8$ can be interpreted as an arbitrage portfolio which bets on large cap versus small cap equities whereas the principal portfolio $\mathcal{F}_7$ is an arbitrage portfolio on FX spread. Principal portfolio $\mathcal{F}_{10}$ seems to be an arbitrage portfolio between the two bond indexes. This is the portfolio with the highest expected Sharpe ratio[19] SR_j and with the highest weight β_j in the MVO portfolio[20]. By using mean-variance optimization, we implicitly have a high exposure on this principal portfolio. Indeed, The US HY index has a high negative expected Sharpe ratio (-1.82) whereas EMBI index has a weak positive expected Sharpe ratio (0.12) resulting in a principal portfolio which is short of the US HY index and long of the EMBI index. Since the correlation between the two indexes is about 60% and their volatilities are low, the principal portfolio has also a low volatility and its weight in the MVO portfolio is dramatically high ($\beta_{10} = 484.58\%$). In this case, the performance of the MVO portfolio is strongly dependent on the performance of the tenth principal portfolio.

We have computed the realized performance of these portfolios over 2007. Results are reported in Table 11. The realized volatility of the MVO portfolio is 14.63%, which is above the targeted volatility of 10%. This suggests that the allocation defined at the end of 2006 was probably too optimistic. We also notice that the riskiest portfolio is the tenth principal portfolio. This result is not surprising because, even if this portfolio was the less risky portfolio in an ex-ante viewpoint, it was also the most leveraged portfolio. In Figure 7, we have represented the cumulative performance of the MVO and $\mathcal{F}_{10}$ portfolios. We can see that their behavior is very close[21]. If we consider the Sharpe ratio, the better portfolio is the principal portfolio $\mathcal{F}_4$, and not the principal portfolio $\mathcal{F}_{10}$, even if the investor has the right views on expected returns[22].

[19] It is equal to 2.58.

[20] The weight of the principal portfolio $\mathcal{F}_{10}$ is equal to 484.58%.

[21] The correlation between the MVO and $\mathcal{F}_{10}$ (resp. $\mathcal{F}_4$) portfolios is 73% (resp. 21%) in 2007.

[22] We remind that the expected returns are exactly equal to the realized returns in 2007.

Table 11: Performance of MVO and principal portfolios in 2007

	MVO	$\mathcal{F}_1$	$\mathcal{F}_2$	$\mathcal{F}_3$	$\mathcal{F}_4$	$\mathcal{F}_5$	$\mathcal{F}_6$	$\mathcal{F}_7$	$\mathcal{F}_8$	$\mathcal{F}_9$	$\mathcal{F}_{10}$
μ_j	50.32	1.60	2.63	0.53	10.04	0.01	3.46	0.98	3.21	3.79	16.27
σ_j	14.63	2.16	2.49	1.29	4.58	0.17	3.01	2.08	5.17	3.48	12.70
SR_j	3.44	0.74	1.05	0.41	2.19	0.06	1.15	0.47	0.62	1.09	1.28

Figure 7: Cumulative performance of MVO and principal portfolios in 2007

Remark 7 *As mentioned earlier, the problem comes from the fact that the MVO portfolio is sensitive to the information matrix. Let us now consider the risk budgeting construction, which is not sensitive to the information matrix but to the covariance matrix. We obtain the results given in Table 12 in the case of the ERC portfolio*[23]*. We can see that the risk allocated to the first principal portfolio is the higher. We retrieve the fact that risk budgeting portfolios make less active bets than MVO portfolios (Roncalli, 2013).*

Table 12: Decomposition of the ERC portfolio at the end of 2006

	$\mathcal{F}_1$	$\mathcal{F}_2$	$\mathcal{F}_3$	$\mathcal{F}_4$	$\mathcal{F}_5$	$\mathcal{F}_6$	$\mathcal{F}_7$	$\mathcal{F}_8$	$\mathcal{F}_9$	$\mathcal{F}_{10}$
β_j	32.70	-0.18	5.46	15.90	-16.83	-32.30	1.02	15.16	-49.36	-87.07
μ_j	2.48	-0.01	0.18	0.24	-0.19	-0.22	0.00	0.02	-0.05	-0.01
σ_j	27.51	20.63	18.32	12.25	10.48	8.30	4.98	3.68	3.16	1.27
SR_j	0.09	0.00	0.01	0.02	-0.02	-0.03	0.00	0.01	-0.02	-0.01

[23] The implied expected returns are then equal to $\mu = \phi \hat{\Sigma} x_{\text{erc}}$ (Roncalli, 2013). The value of ϕ is scaled to target a 10% volatility.

4.2 Regularized portfolios

We continue the previous example by adding a L_2 penalization on the MVO optimization. We report the solution in Table 13 when λ takes the value 0.1%. Even if it is a low penalty, it has a real impact on the allocation. For instance, we notice a high reduction of the US HY exposure. By using this constraint, we modify the composition and the risk associated to each principal portfolio (see Table 14). It helps then to reduce the highest exposures on arbitrage portfolios and to have a more balanced allocation. In our case, we also notice that this is the fourth principal portfolio which now has the highest expected Sharpe ratio (1.94).

Table 13: Comparison of MVO and L_2 portfolios at the end of 2006

Index	MVO	L_2-MVO
SPX	65.57%	51.11%
SX5E	−1.97%	−4.53%
TPX	−24.39%	−38.69%
RTY	−49.54%	−45.53%
MSCI EM	55.60%	66.64%
US HY	−489.99%	−112.58%
EMBI	29.26%	−37.02%
EUR/USD	41.91%	33.25%
JPY/USD	−20.96%	−25.71%
GSCI	2.62%	10.25%

Table 14: Decomposition of the L_2-MVO portfolio at the end of 2006

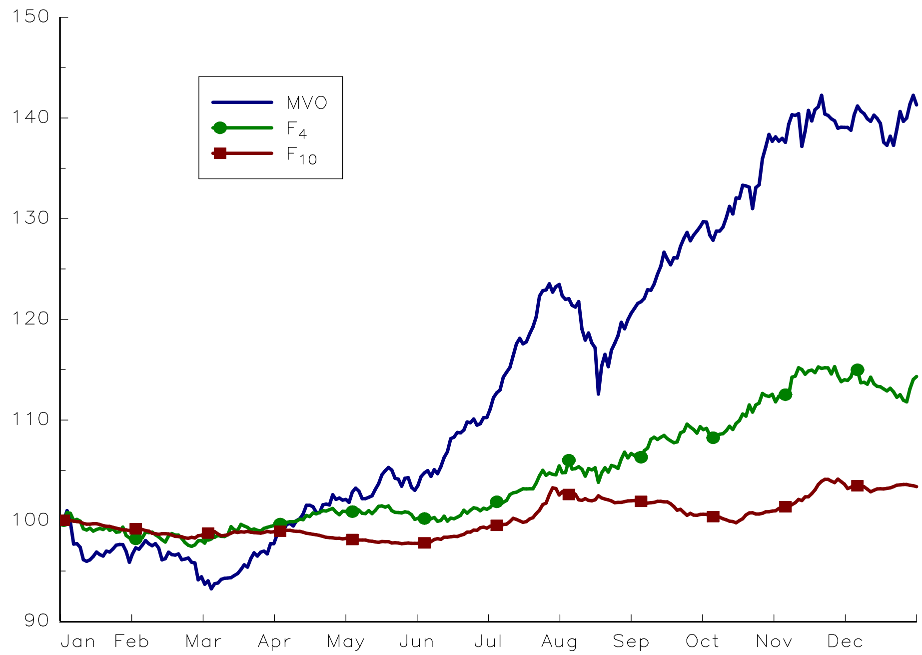
	$\mathcal{F}_1$	$\mathcal{F}_2$	$\mathcal{F}_3$	$\mathcal{F}_4$	$\mathcal{F}_5$	$\mathcal{F}_6$	$\mathcal{F}_7$	$\mathcal{F}_8$	$\mathcal{F}_9$	$\mathcal{F}_{10}$
β_j	10.56	17.81	8.97	55.23	2.06	45.43	33.24	65.66	71.65	101.90
μ_j	22.47	21.53	8.61	24.52	0.69	9.94	3.21	4.29	3.98	3.29
σ_j	27.69	20.87	18.59	12.65	10.95	8.88	5.90	4.85	4.47	3.41
SR_j	0.81	1.03	0.46	1.94	0.06	1.12	0.54	0.88	0.89	0.96

The performance of these allocations is reported in Table 15. We notice that the realized volatility is equal to 11.62%, which is close to the target volatility and that the Sharpe ratio is similar than the one obtained in the non-regularized portfolio. The risk of principal portfolios is now more balanced. These portfolios benefit then from regularization. Indeed, the correlation of the L_2-MVO portfolio with the principal portfolio $\mathcal{F}_{10}$ is now equal to −2% whereas the correlation with the principal portfolio $\mathcal{F}_4$ is 47%.

Table 15: Performance of L_2-MVO and principal portfolios in 2007

	L_2-MVO	$\mathcal{F}_1$	$\mathcal{F}_2$	$\mathcal{F}_3$	$\mathcal{F}_4$	$\mathcal{F}_5$	$\mathcal{F}_6$	$\mathcal{F}_7$	$\mathcal{F}_8$	$\mathcal{F}_9$	$\mathcal{F}_{10}$
μ_j	41.37	2.37	3.87	0.77	14.37	0.01	4.55	1.05	2.79	2.85	3.36
σ_j	11.62	3.21	3.67	1.89	6.45	0.23	3.95	2.22	4.47	2.62	2.67
SR_j	3.56	0.74	1.05	0.41	2.23	0.06	1.15	0.47	0.62	1.09	1.26

Figure 8: Cumulative performance of L_2-MVO and principal portfolios in 2007



Remark 8 *We modify the previous example by considering unperfect views. Suppose that we have wrong expected returns for the bond indexes, i.e.* 3.07% *for the US HY index and* -0.5% *for the EMBI index. In this case, exposures change slightly except for the principal portfolio* $\mathcal{F}_{10}$*, which becomes* -452%*. We have reported in Table 16 the performance of the optimized and principal portfolios. We can see that having the wrong views has a big impact on the MVO portfolio. Its return is reduced by* 81%*! For the* L_2-*MVO portfolio, the return only decreases by* 29%*.*

Table 16: Performance of optimized and principal portfolios in 2007 with wrong views

	L_2-MVO	$\mathcal{F}_1$	$\mathcal{F}_2$	$\mathcal{F}_3$	$\mathcal{F}_4$	$\mathcal{F}_5$	$\mathcal{F}_6$	$\mathcal{F}_7$	$\mathcal{F}_8$	$\mathcal{F}_9$	$\mathcal{F}_{10}$
						MVO					
μ_j	11.19	1.69	2.76	0.55	10.64	0.01	3.60	1.05	3.42	3.56	-14.38
σ_j	17.57	2.28	2.62	1.35	4.84	0.17	3.13	2.24	5.51	3.27	11.85
SR_j	0.64	0.74	1.05	0.41	2.20	0.06	1.15	0.47	0.62	1.09	-1.21
						L_2-MVO					
μ_j	32.43	2.39	3.88	0.76	14.53	0.01	4.52	1.08	2.84	2.55	-2.96
σ_j	12.91	3.23	3.68	1.88	6.52	0.22	3.93	2.29	4.56	2.35	2.38
SR_j	2.51	0.74	1.05	0.41	2.23	0.06	1.15	0.47	0.62	1.09	-1.24

4.3 Hedging portfolios

We recall that the MVO portfolio involves the inverse of the covariance matrix. Since the information matrix reflects the hedging relationships, we illustrate how penalization methods may be used to obtain better hedging portfolios. For that, we impose a L_1 constraint on hedging regressions in order to build sparse hedging portfolios. To choose the penalization parameter λ, we use the BIC criterion as described in Zou *et al.* (2007).

We consider the universe of the 10 diversified indexes and compute the hedging portfolios based on the daily asset returns for 2006. Results are reported in Tables 17 and 18. For each asset i, we indicate the hedging coefficients $\hat{\beta}_i$, the standard deviation $\hat{s}_i$ of the residuals and the coefficient of determination R_i^2. These statistics are expressed in %. We notice that OLS regression produces noisy portfolios with very small exposures on some assets. For instance, the weight of the Topix index is 0.4% in the S&P 500 hedging portfolio. With the Lasso regression, the hedging portfolio is more sparse without a high decrease of the coefficient R_i^2. For instance, the S&P 500 index is hedged with the Russell 2000 and Eurostoxx 50 indexes whereas the difference in terms of R^2 is only 1%. The long-short exposure on US HY and EMBI and the short exposure on EM equities vanish. From an economic point of view, the Lasso hedging portfolios is then more reliable than the OLS hedging portfolio.

Table 17: OLS hedging portfolios (in %) at the end of 2006

	SPX	SX5E	TPX	RTY	EM	US HY	EMBI	EUR	JPY	GSCI
SPX		58.6	6.0	150.3	-30.8	-0.5	5.0	-7.3	15.3	-25.5
SX5E	9.0		-1.2	-1.3	35.2	0.8	3.2	-4.5	-5.0	-1.5
TPX	0.4	-0.6		-2.4	38.1	1.1	-3.5	-4.9	-0.8	-0.3
RTY	48.6	-2.7	-10.4		26.2	-0.6	1.9	0.2	-6.4	5.6
EM	-4.1	30.9	69.2	10.9		0.9	4.6	9.1	3.9	33.1
US HY	-5.0	53.5	160.0	-18.8	69.5		95.6	48.4	31.4	-211.7
EMBI	10.8	44.2	-102.1	12.3	73.4	19.4		-5.8	40.5	86.2
EUR	-3.6	-14.7	-33.4	0.3	33.8	2.3	-1.4		56.7	48.2
JPY	6.8	-14.5	-4.8	-8.8	12.7	1.3	8.4	50.4		-33.2
GSCI	-1.1	-0.4	-0.2	0.8	10.7	-0.9	1.8	4.2	-3.3	
$\hat{s}_i$	0.3	0.7	0.9	0.5	0.7	0.1	0.2	0.4	0.4	1.2
R_i^2	83.0	47.7	34.9	82.4	60.9	39.8	51.6	42.3	43.7	12.1

Table 18: Lasso hedging portfolios (in %) at the end of 2006

	SPX	SX5E	TPX	RTY	EM	US HY	EMBI	EUR	JPY	GSCI
SPX		49.2		146.8			5.0	-3.2		
SX5E	5.1				32.3		3.2			
TPX					37.4	0.8	-3.1			
RTY	46.8		-3.1		10.4		1.9			
EM		25.0	61.3	6.5		0.8	4.3	2.6		22.4
US HY			82.2		65.9		93.8	19.1	20.7	
EMBI		24.9	-70.3		71.6	17.5			33.8	
EUR			-23.7		33.6	1.4			51.9	13.8
JPY					9.3	1.1	7.6	48.9		
GSCI					10.2	-0.3	1.6	2.9		
$\hat{s}_i$	0.3	0.7	1.0	0.5	0.7	0.1	0.2	0.4	0.4	1.3
R_i^2	82.0	44.9	33.9	82.1	60.4	38.0	51.5	39.7	41.5	7.8

We now consider the example of the S&P 100 universe for the period January 2000 to

December 2011. We compute the minimum variance portfolio using the hedging relationships:

$$x_i^\star(\gamma) = \gamma \frac{1 - \hat{\beta}_i^\top \mathbf{1}}{\hat{s}_i^2}$$

γ is a scaling parameter such that the sum of weights is equal to 100%. We rebalance the portfolio every month whereas the information matrix is computed using a rolling window of 260 trading days. Results are reported in Table 19. We notice that the Lasso-MV portfolio improves the performance of the OLS-MV portfolio (higher return $\mu(x)$, lower volatility $\sigma(x)$ and lower drawdown $\mathcal{MDD}$). Moreover, we notice that the turnover τ is dramatically reduced thanks to the norm constraint. This result suggests that a lot of hedging relationships made by non-regularized MVO portfolios are not optimal. If we compute the sparsity rate[24] of the information matrix at each rebalancing dates, its range is between 0% and 0.02% for the OLS regression whereas it is between 73.70% and 87.14% for the Lasso regression. We verify that the Lasso-MV portfolio produces sparse hedging relationships, which is not the case with the traditional OLS-MV portfolio. For instance, Google is hedged by 99 stocks at December 2011, if we consider the OLS-MV portfolio. Using the L_1 constraint, Google is hedged by only 13 stocks[25]. In Figure 9, we have reported some statistics about the 100 hedging relationships at December 2011. The coefficient R^2 is reduced when we consider Lasso regression, meaning that Lasso hedging portfolios have a lower in-the-sample explanatory power. Nevertheless, the Lasso net exposure $\sum_{j \neq i} \hat{\beta}_{i,j}$ is generally close to the OLS net exposure. In fact, one of the benefits with Lasso regression is to reduce the short exposure[26] of the hedging portfolio. In the end, Lasso hedging portfolios are less leveraged[27]. In this case, it is obvious that the Lasso-MV portfolio uses a less noisy information matrix than the classical MV portfolio.

Table 19: Performance of OLS-MV and Lasso-MV portfolios

	$\mu(x)$	$\sigma(x)$	$\mathrm{SR}(x)$	$\mathcal{MDD}$	τ
OLS-MV	3.60%	14.39%	0.25	−39.71%	19.4
Lasso-MV	5.00%	13.82%	0.36	−35.42%	5.9

5 Conclusion

In this paper, we have reviewed the different approaches of portfolio allocation. The first generation of methods (resampling, random matrix theory and covariance shrinkage), which were proposed at the end of the Nineties, aim to reduce the noisy part of the covariance matrix. However, this type of approach is not sufficient. The second generation of methods (Lasso and ridge regression) seeks to directly regularize the MVO portfolio by introducing more sparsity in the solution. These sparse methods are more satisfactory because they produce more robust portfolios that are less sensitive to input parameters.

The first application of the paper clearly shows how the noisy part of the covariance matrix impacts the optimized portfolio. Principal portfolios are then the adequate tool

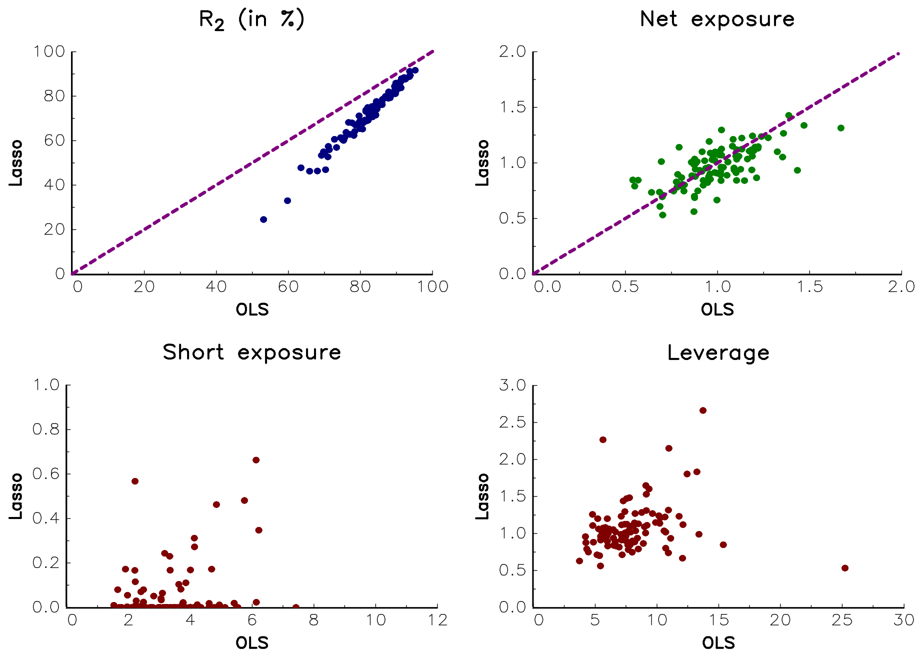
[24] We consider the sparsity measure ℓ_ϵ^0 defined in Hurley and Rickard (2009) with $\epsilon = 10^{-5}$.

[25] They are Boeing (4.6%), United technologies (1.1%), Schlumberger (1.8%), Williams cos. (1.8%), Microsoft (13.7%), Honeywell intl. (2.7%), Caterpillar (0.9%), Apple (25.0%), Mastercard (2.5%), Devon energy (2.9%), Nike (1.2%), Amazon (6.7%) and Apache (8.7%).

[26] It is measured as the opposite of $\sum_{j \neq i} \min\left(0, \hat{\beta}_{i,j}\right)$.

[27] We compute the leverage as $\sum_{j \neq i} \left|\hat{\beta}_{i,j}\right|$.

Figure 9: Statistics of OLS and Lasso hedging relationships (December 2011)



to understand the relationships between the MVO portfolio and the eigendecomposition of the covariance matrix. The second application illustrates how penalization methods make it possible to obtain portfolios that are less sensitive to the noise of the covariance matrix. With this type of regularization, investment bets are less aggressive and optimized portfolios are more robust. Finally, the last application proposes to interpret mean-variance optimization as an allocation model that takes exposures with respect to hedging portfolios. In this case, the regularization method consists of introducing robustness in the estimation of hedging portfolios. By using sparse regression methods, portfolio turnover is dramatically reduced and robustness is improved.

The improvement of the Markowitz model is an endless issue. The most common proposed solutions are generally less than satisfactory. This is the case of the three main approaches: resampling, denoising and shrinkage methods. Other sophisticated methods exist, like the robust approach of Tütüncü and Koenig (2004), but they are not used by practitioners because they failed to significantly improve Markowitz portfolios. A new form of portfolio regularization has been introduced recently by Brodie *et al* (2009) and DeMiguel *et al* (2009). Contrary to the previous approaches, these methods of sparse portfolio allocation aim to directly regularize the solution instead of the input parameters. In particular, Lasso and ridge methods are today largely used by portfolio managers. They have helped to rehabilitate the Markowitz model in the active management field when the goal is precisely to incorporate views or bets in actively managed long-short portfolios.

A Mathematical results

A.1 The single-factor model

We assume that the first factor of the covariance matrix corresponds to the market factor and that the idiosyncratic volatilities are equal ($\tilde{\sigma}_1 = \ldots = \tilde{\sigma}_n = \tilde{\sigma}$). In terms of eigenvectors, we have:

$$v_1 = \frac{\beta}{\sqrt{\sum_{i=1}^n \beta_i^2}}$$

Because $VV^\top = I$ and $\lambda_1 \simeq \sigma_m^2 \sum_{i=1}^n \beta_i^2 + \tilde{\sigma}^2$, we obtain the following expression of $\tilde{\Sigma}_\alpha$:

$$\begin{aligned}
\tilde{\Sigma}_\alpha &= \alpha \hat{\Phi} + (1-\alpha)\,\hat{\Sigma} \\
&= \alpha \left(\sigma_m^2 \beta\beta^\top + D\right) + (1-\alpha)\, V \Lambda V^\top \\
&= \alpha \left(\sigma_m^2 \beta\beta^\top + DVV^\top\right) + (1-\alpha) \sum_{i=1}^n \lambda_i v_i v_i^\top \\
&\simeq \alpha \sigma_m^2 \beta\beta^\top + \alpha DVV^\top + (1-\alpha)\,\lambda_1 \frac{\beta\beta^\top}{\sum_{i=1}^n \beta_i^2} + (1-\alpha) \sum_{i=2}^n \lambda_i v_i v_i^\top \\
&= \left(\alpha \sigma_m^2 + (1-\alpha) \frac{\lambda_1}{\sum_{i=1}^n \beta_i^2}\right) \beta\beta^\top + (1-\alpha) \sum_{i=2}^n \lambda_i v_i v_i^\top + \alpha DVV^\top
\end{aligned}$$

It follows that:

$$\begin{aligned}
\tilde{\Sigma}_\alpha &\simeq \left(\alpha \sigma_m^2 + (1-\alpha) \frac{\lambda_1}{\sum_{i=1}^n \beta_i^2}\right) \beta\beta^\top + (1-\alpha) \sum_{i=2}^n \lambda_i v_i v_i^\top + \\
&\quad\; \alpha \left(\tilde{\sigma}^2 \frac{\beta\beta^\top}{\sum_{i=1}^n \beta_i^2} + \sum_{i=2}^n \tilde{\sigma}^2 v_i v_i^\top\right) \\
&= \left(\alpha \sigma_m^2 + (1-\alpha) \frac{\lambda_1}{\sum_{i=1}^n \beta_i^2} + \alpha \frac{\tilde{\sigma}^2}{\sum_{i=1}^n \beta_i^2}\right) \beta\beta^\top + \\
&\quad\; (1-\alpha) \sum_{i=2}^n \lambda_i v_i v_i^\top + \alpha \sum_{i=2}^n \tilde{\sigma}^2 v_i v_i^\top \\
&= \lambda_1 \frac{\beta\beta^\top}{\sum_{i=1}^n \beta_i^2} + \sum_{i=2}^n \left((1-\alpha)\,\lambda_i + \alpha\tilde{\sigma}^2\right) v_i v_i^\top \\
&= \lambda_1 v_1 v_1^\top + \sum_{i=2}^n \left((1-\alpha)\,\lambda_i + \alpha\tilde{\sigma}^2\right) v_i v_i^\top
\end{aligned}$$

A.2 Analytical solutions of L_2 portfolio optimization

A.2.1 The case of variance penalization

We have the following optimization program:

$$x^\star(\gamma, \lambda) = \arg\min \frac{1}{2} x^\top \left(\hat{\Sigma} + \lambda A\right) x - \gamma x^\top \hat{\mu}$$

We assume that A is the covariance matrix without correlations. We have $A = \upsilon = \operatorname{diag}\left(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2\right)$. It follows that:

$$\begin{aligned}
\tilde{\Sigma} &= \hat{\Sigma} + \lambda \upsilon \\
&= \upsilon^{1/2} \hat{C} \upsilon^{1/2} + \lambda \upsilon^{1/2} \upsilon^{1/2} \\
&= \upsilon^{1/2} \left(\hat{C} + \lambda I_n\right) \upsilon^{1/2}
\end{aligned}$$

By setting $y = \upsilon^{1/2} x$, we get:

$$y^\star(\gamma, \lambda) = \arg\min \frac{1}{2} y^\top \left(\hat{C} + \lambda I_n\right) y - \gamma y^\top \hat{s}$$

where $\hat{s}$ is the vector of expected Sharpe ratios. The solution is then:

$$\begin{aligned}
x^\star(\gamma, \lambda) &= \upsilon^{-1/2} y^\star(\gamma, \lambda) \\
&= \gamma \upsilon^{-1/2} \left(\hat{C} + \lambda I_n\right)^{-1} \hat{s}
\end{aligned}$$

When λ tends to ∞ and when we renormalize the solution, we retrieve the analytical expression of Merton (1969):

$$\lim_{\lambda \to \infty} x_i^\star(\gamma, \lambda) \propto \gamma \frac{\hat{\mu}_i}{\hat{\sigma}_i^2}$$

A.2.2 The case of a target portfolio

We remind that:

$$x^\star(\gamma, \lambda) = \left(\hat{\Sigma} + \lambda I_n\right)^{-1} \left(\gamma \hat{\mu} + \lambda x_0\right)$$

We see that:

$$\begin{aligned}
I_n - \left(I_n + \lambda \hat{\Sigma}^{-1}\right)^{-1} &= \left(I_n + \lambda \hat{\Sigma}^{-1}\right)^{-1} \left(I_n + \lambda \hat{\Sigma}^{-1}\right) - \left(I_n + \lambda \hat{\Sigma}^{-1}\right)^{-1} \\
&= \left(I_n + \lambda \hat{\Sigma}^{-1}\right)^{-1} \lambda \hat{\Sigma}^{-1}
\end{aligned}$$

We then obtain:

$$\begin{aligned}
x^\star(\gamma, \lambda) &= \left(\hat{\Sigma} + \lambda I_n\right)^{-1} \gamma \hat{\mu} + \left(\hat{\Sigma} + \lambda I_n\right)^{-1} \lambda x_0 \\
&= \left(I_n + \lambda \hat{\Sigma}^{-1}\right)^{-1} x^\star(\gamma) + \left(I_n + \lambda \hat{\Sigma}^{-1}\right)^{-1} \lambda \hat{\Sigma}^{-1} x_0 \\
&= B x^\star(\gamma) + \left(I_n - B\right) x_0
\end{aligned}$$

where $B = \left(I_n + \lambda \hat{\Sigma}^{-1}\right)^{-1}$.

A.3 Relationship between penalization and robust portfolio optimization

A.3.1 Robust portfolio optimization

Robust optimization is a technique designed to build a portfolio that performs well in a number of different scenarios including the extreme ones. For instance, a portfolio manager may apply a confidence interval on each coefficients of the estimated covariance matrix. Halldórsson and Tütüncü (2003) have extensively studied this problem and define the uncertainty set $\mathcal{U}$ as follows:

$$\mathcal{U} = \left\{\Sigma : \Sigma^- \leq \Sigma \leq \Sigma^+, \Sigma \succeq 0\right\}$$

where Σ^- and Σ^+ are extreme values of the set $\mathcal{U}$. In the case of the MVO problem, we obtain:

$$x^\star(\gamma; \mathcal{U}) = \arg\min \max_{\Sigma \in \mathcal{U}} \frac{1}{2} x^\top \Sigma x - \gamma x^\top \hat{\mu}$$

Under the constraint $x \geq \mathbf{0}$, we notice that the solution is:

$$x^\star(\gamma; \mathcal{U}) = \gamma \left(\Sigma^+\right)^{-1} \hat{\mu}$$

A.3.2 QP formulation of the optimization problem

By decomposing the weights as $x_i = x_i^+ - x_i^-$ with $x_i^+ \geq 0$ and $x_i^- \geq 0$, we obtain:

$$\begin{aligned}
\frac{1}{2} x^\top \Sigma x - \gamma x^\top \hat{\mu} &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \Sigma_{i,j} x_i x_j - \gamma x^\top \hat{\mu} \\
&= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \Sigma_{i,j} x_i^+ x_j^+ + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \Sigma_{i,j} x_i^- x_j^- \\
&\quad - \sum_{i=1}^n \sum_{j=1}^n \Sigma_{i,j} x_i^+ x_j^- - \gamma x^\top \hat{\mu}
\end{aligned}$$

Let us consider the following function:

$$m(x) = \max_{\Sigma \in \mathcal{U}} \frac{1}{2} x^\top \Sigma x - \gamma x^\top \hat{\mu}$$

For a given set of weights x, the worst covariance matrix among $\mathcal{U}$ is the highest (resp. the lowest) element of $\mathcal{U}$ if $x_i x_j \geq 0$ (resp. $x_i x_j \leq 0$). It comes that:

$$m(x) = \frac{1}{2} \tilde{x}^\top \tilde{\Sigma} \tilde{x} - \gamma \tilde{x}^\top \tilde{\mu}$$

with:

$$\tilde{x} = \begin{pmatrix} x^+ \\ x^- \end{pmatrix}, \quad \tilde{\Sigma} = \begin{pmatrix} \Sigma^+ & -\Sigma^- \\ -\Sigma^- & \Sigma^+ \end{pmatrix} \quad \text{and} \quad \tilde{\mu} = \begin{pmatrix} \hat{\mu} \\ -\hat{\mu} \end{pmatrix}$$

We finally obtain:

$$\begin{aligned}
x^\star(\gamma; \mathcal{U}) &= \arg\min \frac{1}{2} \tilde{x}^\top \tilde{\Sigma} \tilde{x} - \gamma \tilde{x}^\top \tilde{\mu} \\
&\text{u.c.} \quad \tilde{x} \geq \mathbf{0}
\end{aligned}$$

This problem can be solved using classical QP algorithm under the condition that $\tilde{\Sigma}$ is positive definite.

A.3.3 L_1 formulation of the optimization problem

Robust estimation is not frequently used in practice, because it is time consuming and it is extremely difficult to define the set $\mathcal{U}$. If we omit the definite positive condition $\Sigma \succeq 0$, we can nevertheless use the L_1 approach to solve a similar problem with the set $\mathcal{U}$ defined as follows:

$$\mathcal{U} = \left\{ \Sigma : \hat{\Sigma} - A \leq \Sigma \leq \hat{\Sigma} + A \right\}$$

where A is a positive definite matrix of noise. We have[28]:

$$\begin{aligned}
\tilde{x}^\top \tilde{\Sigma} \tilde{x} &= \tilde{x}^\top \begin{pmatrix} \hat{\Sigma} + A & -\hat{\Sigma} + A \\ -\hat{\Sigma} + A & \hat{\Sigma} + A \end{pmatrix} \tilde{x} \\
&= \tilde{x}^\top \begin{pmatrix} \hat{\Sigma} & -\hat{\Sigma} \\ -\hat{\Sigma} & \hat{\Sigma} \end{pmatrix} \tilde{x} + \tilde{x}^\top \begin{pmatrix} A & A \\ A & A \end{pmatrix} \tilde{x} \\
&= \left(x^+ - x^- \right)^\top \hat{\Sigma} \left(x^+ - x^- \right) + \left(x^+ + x^- \right)^\top A \left(x^+ + x^- \right) \\
&= x^\top \hat{\Sigma} x + \left| x \right|^\top A \left| x \right|
\end{aligned}$$

We deduce that the minmax problem can be reformulated as follows:

$$x^\star \left(\gamma ; A \right) = \arg\min \frac{1}{2} x^\top \hat{\Sigma} x - \gamma x^\top \hat{\mu} + \frac{1}{2} \left| x \right|^\top A \left| x \right| \tag{5}$$

Using a confidence interval on the empirical covariance matrix implies then a penalization on the squared L_1 norm.

Remark 9 *If $A = cI_n$ with c a scalar, the optimization problem (5) is a L_2 constrained problem. In this case, we have:*

$$\mathcal{U} = \left\{ \Sigma : V \left(\Lambda - A \right) V^\top \leq \Sigma \leq V \left(\Lambda + A \right) V^\top \right\}$$

where $V \Lambda V^\top$ is the eigendecomposition of $\hat{\Sigma}$. As a result, the L_2 constraint is also a minmax problem with uncertain eigenvalues. This gives some intuitions to choose the L_2 shrinkage parameter λ equal to the average of diagonal elements of A.

A.4 Relationship between the conditional normal distribution and the linear regression

Let us consider a Gaussian random vector defined as follows:

$$\begin{pmatrix} Y \\ X \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix}, \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix} \right)$$

The conditional distribution of Y given $X = x$ is a multivariate normal distribution. We have (Roncalli, 2013):

$$\begin{aligned}
\mu_{y|x} &= \mathbb{E} \left[Y \mid X = x \right] \\
&= \mu_y + \Sigma_{yx} \Sigma_{xx}^{-1} \left(x - \mu_x \right)
\end{aligned}$$

and:

$$\begin{aligned}
\Sigma_{yy|x} &= \sigma^2 \left[Y \mid X = x \right] \\
&= \Sigma_{yy} - \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy}
\end{aligned}$$

We deduce that:

$$Y = \mu_y + \Sigma_{yx} \Sigma_{xx}^{-1} \left(x - \mu_x \right) + u$$

[28]Since $\hat{\Sigma}$ and A are two positive definite matrices, $\tilde{\Sigma}$ is also a positive definite matrix.

where u is a centered Gaussian random variable with variance $s^2 = \Sigma_{yy|x}$. It follows that:

$$Y = \underbrace{\left(\mu_y - \Sigma_{yx}\Sigma_{xx}^{-1}\mu_x\right)}_{\beta_0} + \underbrace{\Sigma_{yx}\Sigma_{xx}^{-1}}_{\beta^\top}x + u$$

We recognize the linear regression of Y on X:

$$Y = \beta_0 + \beta^\top x + u$$

Moreover, we have:

$$\begin{aligned} R^2 &= 1 - \frac{s^2}{\Sigma_{yy}} \\ &= \frac{\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}}{\Sigma_{yy}} \end{aligned}$$

References

[1] Black F. and Litterman R.B. (1992), Global Portfolio Optimization, *Financial Analysts Journal*, 48(5), pp. 28-43.

[2] Broadie M. (1993), Computing Efficient Frontiers using Estimated Parameters, *Annals of Operations Research*, 45(1), pp. 21-58.

[3] Brodie J., Daubechies I., De Mol C., Giannone D. and Loris I. (2009), Sparse and Stable Markowitz Portfolios, *Proceedings of the National Academy of Sciences*, 106(30), pp. 12267-12272.

[4] Candelon B., Hurlin C. and Tokpavi S. (2012), Sampling Error and Double Shrinkage Estimation of Minimum Variance Portfolios, *Journal of Empirical Finance*, 19(4), pp. 511-527.

[5] DeMiguel V., Garlappi L., Nogales F.J. and Uppal R. (2009), A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms, *Management Science*, 55(5), pp. 798-812.

[6] DeMiguel V., Garlappi L. and Uppal R. (2009), Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?, *Review of Financial Studies*, 22(5), pp. 1915-1953.

[7] DeMiguel V., Martin-Utrera A. and Nogales F.J. (2011), Size Matters: Optimal Calibration of Shrinkage Estimators for Portfolio Selection, *SSRN*, `www.ssrn.com/abstract=1891847`.

[8] Dupleich Ulloa M.R., Giamouridis D. and Montagu C. (2012), Risk Reduction in Style Rotation, *Journal of Portfolio Management*, 38(2), pp. 44-55.

[9] El Ghaoui L., Oks M. and Oustry F. (2003), Worst-Case Value-at-Risk and Robust Portfolio Optimization: A Conic Programming Approach, *Operations Research*, 51(4), pp. 543-556.

[10] Halldórsson B.V. and Tütüncü R.H. (2003), An Interior-Point Method for a Class of Saddle-Point Problems, *Journal of Optimization Theory and Applications*, 116(3), pp. 559-590.

[11] Hastie T., Tibshirani R. and Friedman J. (2009), *The Elements of Statistical Learning*, Second Edition. Springer.

[12] Hurley N. and Rickard S. (2009), Comparing Measures of Sparsity, *IEEE Transactions on Information Theory*, 55(10), pp. 4723-4741.

[13] Jagannathan R. and Ma T. (2003), Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps, *Journal of Finance*, 58(4), pp. 1651-1684.

[14] James G. and Hastie T. (1997), Generalizations of the Bias/Variance Decomposition for Prediction Error, *Technical Report*, Department of Statistics, Stanford University.

[15] Jorion P. (1992), Portfolio Optimization in Practice, *Financial Analysts Journal*, 48(1), pp. 68-74.

[16] Laloux L., Cizeau P., Bouchaud J-P. and Potters M. (1999), Noise Dressing of Financial Correlation Matrices, *Physical Review Letters*, 83(7), pp. 1467-1470.

[17] Ledoit O. and Wolf M. (2003), Improved Estimation of the Covariance Matrix of Stock Returns With an Application to Portfolio Selection, *Journal of Empirical Finance*, 10(5), pp. 603-621.

[18] Ledoit O. and Wolf M. (2004), A Well-conditioned Estimator for Large-dimensional Covariance Matrices, *Journal of Multivariate Analysis*, 88(2), pp. 365-411.

[19] Ledoit O. and Wolf M. (2004), Honey, I Shrunk the Sample Covariance Matrix, *Journal of Portfolio Management*, 30(4), pp. 110-119.

[20] Marchenko V.A. and Pastur L.A. (1967), Distribution of Eigenvalues for Some Sets of Random Matrices, *Mathematics of the USSR-Sbornik*, 72(4), pp. 507-536.

[21] Markowitz H. (1952), Portfolio Selection, *Journal of Finance*, 7(1), pp. 77-91.

[22] Michaud R.O. (1989), The Markowitz Optimization Enigma: Is 'Optimized' Optimal?, *Financial Analysts Journal*, 45(1), pp. 31-42.

[23] Michaud R.O. (1998), *Efficient Asset Management: A Practical Guide to Stock Portfolio Management and Asset Allocation*, Financial Management Association, Survey and Synthesis Series, HBS Press.

[24] Merton R.C. (1969), Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case, *Review of Economics and Statistics*, 51(3), pp. 247-257.

[25] Roncalli T. (2013), *Introduction to Risk Parity and Budgeting*, Chapman & Hall/CRC Financial Mathematics Series.

[26] Scherer B. (2002), Portfolio Resampling: Review and Critique, *Financial Analysts Journal*, 58(6), pp. 98-109.

[27] Scherer B. (2007), *Portfolio Construction & Risk Budgeting*, Third edition, Risk Books.

[28] Sharpe W.F. (1963), A Simplified Model for Portfolio Analysis, *Management Science*, 9(2), pp. 277-293.

[29] Stevens G.V.G. (1998), On the Inverse of the Covariance Matrix in Portfolio analysis, *Journal of Finance*, 53(5), pp. 1821-1827.

[30] Tibshirani R. (1996), Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society B*, 58(1), pp. 267-288.

[31] Tu J. and Zhou G. (2011), Markowitz meets Talmud: A Combination of Sophisticated and Naive Diversification Strategies, *Journal of Financial Economics*, 99(1), pp. 204-215.

[32] Tütüncü R.H. and Koenig M. (2004), Robust Asset Allocation, *Annals of Operations Research*, 132(1-4), pp. 157-187.

[33] Zou H. and Hastie T. (2005), Regularization and Variable Selection via the Elastic Net, *Journal of the Royal Statistical Society B*, 67(2), pp. 301-320.

[34] Zou H., Hastie T. and Tibshirani R. (2007), On the "Degrees of Freedom" of the Lasso, *Annals of Statistics*, 35(5), pp. 2173-2192.